



## **Aid for Trade: A Meta-evaluation**

**Claire Delpeuch, Marie-Agnès Jouanjean, Alexandre Le Vernoy,  
Patrick Messerlin, Thomas Orliac**



## TABLE OF CONTENTS

EXECUTIVE SUMMARY .....	5
INTRODUCTION .....	7
CHAPTER 1. PURPOSES, SCOPE AND METHODOLOGY .....	9
Section 1. Two purposes .....	9
Section 2. Scope .....	9
Section 3. Methodology .....	10
CHAPTER 2. DO THE EVALUATIONS DELIVER WHAT WE WANT TO KNOW? .....	11
Summary of the results.....	11
Introduction .....	11
Section 1. The trade component.....	12
Section 2. The development component .....	20
Section 3. Concluding remarks .....	27
CHAPTER 3: QUALITATIVE ANALYSIS .....	29
Summary of the results.....	29
Introduction .....	29
Section 1. Aid for Trade in Vietnam.....	31
Section 2. Aid for Trade in Ghana .....	36
Section 3. Aid for Trade in TranStor.....	38
Section 4. Final remarks.....	42
CHAPTER 4. WHAT NEEDS TO BE DONE: LOOKING FOR A SIMPLE APPROACH ...	43
Summary of the results.....	43
Introduction .....	43
Section 1. Preparing the ground .....	43
Section 2. Options for a better assessment of the Aid for Trade operations .....	45
ANNEX A. AN OVERVIEW OF THE EVALUATIONS, EVALUATORS AND THE TIME PATTERN.....	48
Section 1. The stock of evaluations under scrutiny .....	48
Section 2. Different types of operations.....	49
Section 3. The time pattern of the evaluations by recipient .....	50
Section 4. The evaluations by donor .....	52
ANNEX B. A BRIEF PRESENTATION OF THE LIST OF KEY WORDS.....	55
BIBLIOGRAPHY.....	57

## Tables

Table 2.1A. The role of trade by recipient .....	14
Table 2.1B. The role of trade by donor, the overall set of evaluations .....	15
Table 2.2A. The role of trade policy, by recipient .....	16
Table 2.2B. The role of trade policy, by donor .....	17
Table 2.3A. The Aid for Trade wide set of trade issues, by recipient .....	18
Table 2.3B. The Aid for Trade wide set of trade issues, by donor .....	19
Table 2.4A. International trade agreements, by recipient .....	20
Table 2.4B. International trade agreements, by donor .....	20
Table 2.5A. Public sector-related constraints, by recipient .....	21
Table 2.5B. Public sector-related constraints, by donor .....	23
Table 2.6A. Supply-side constraints, by recipient .....	24
Table 2.6B. Supply-side constraints, by donor .....	25
Table 2.7A. Development performances, by recipient .....	26
Table 2.7B. Development performances, by donor .....	27
Table 3.1. Key data for the narrow set of evaluations: Vietnam .....	31
Table 3.2. Key data for the narrow set of evaluations: the TranStor sector .....	39
Table 4.1. Key evaluation procedures, by recipient .....	44
Table 4.2. Key econometric techniques, by recipient .....	44
Table 4.3. Key econometric techniques, by recipient .....	45
Table 4.4 Impact assessment: an ideal checklist .....	46
Table A1. An overview: the stock of evaluations (1999-2009) .....	48
Table A2. A breakdown by project, programme and general aid policy .....	50
Table A3. An overview of the overall set of evaluations: the time pattern by recipient .....	51
Table A4. An overview of the narrow set of evaluations: the time pattern by recipient .....	52
Table A5. An overview of the overall set of evaluations: the donor pattern .....	53
Table A6. An overview of the narrow set of evaluations: the donor pattern .....	54
Table B1. The complete list of words and expressions .....	55
Table B2. The core list of words and expressions key from an economic perspective .....	56

## Figures

Figure 2.1: Aid for Trade: the agenda .....	12
---	----

## EXECUTIVE SUMMARY

This following study is a “meta-evaluation” of the way aid agencies evaluate and assess their trade-related operations – and was undertaken as part of the OECD’s broader efforts to improve the effectiveness of aid-for-trade. It focuses on 162 evaluations of operations in two countries - Ghana and Vietnam - and two sectors - transport and storage - between 1999 and 2010. In particular, it looks at whether trade was a true objective of the operations under scrutiny, and whether trade and development outcomes were evaluated. More broadly, it asks whether the evaluations selected offer the information that policy makers in charge of international aid want – and need - to get from field evaluators.

This report's methodology rests on quantitative measurement as well as qualitative analysis. It examines how often key words or expressions occur in the entire set of evaluations, in an effort to reveal the implicit interests of evaluators. It does the same with a narrower set of evaluations that are much more closely focused on trade-related matters. Next it undertakes a closer reading of a more narrow set of evaluations, looking beyond the immediate issues treated by the evaluators and analyzing the content of key words and phrases in order to determine how core issues were considered.

This dual approach leads to the following conclusions. First, evaluations of aid-for-trade operations do not say much about trade. “Trade” and “exports” were not among the most frequently mentioned words, while “imports” was almost completely ignored. Similarly, references to the World Trade Organization (WTO) or regional trade agreements were largely absent from the evaluations. The qualitative reading yields similar conclusions. Not only was the trade impact of operations clearly not the main focus of evaluators’ work, but — in a number of cases — it is not even addressed.

Second, and unexpectedly, the evaluations usually did not clarify the policy linkages which matter most to policy makers. For example, the potential impact—positive or negative—of trade policy instruments, such as tariffs, quotas or subsidies, on funded projects was not assessed or even examined. Nor were the linkages with “behind the border” measures, such as regulatory reforms or private-sector policies, evaluated. While some relevant terms appear in a number of evaluations, subsequent qualitative analysis suggests that these terms were not well understood or defined by evaluators, and their usage was not substantiated by reference to hard evidence.

Third, in sharp contrast with the surprising silence on trade-related issues, the evaluations referred extensively to broad, development-related concepts, such as gender or poverty reduction, but without clearly defining these terms. This tendency to favour generic concepts over precise terms often meant that the evaluations were vague and ill-focused. For instance, gender was either mentioned in passing, as a cross-cutting issue in operations, or measured in terms of short-term objectives that had little tangible relevance to a project’s impact on women’s economic or societal situation (e.g., they merely added up the number of women participating in a given seminar).

Fourth, the evaluations often lacked an adequate or realistic timeframe for measuring results, rarely distinguishing between what was achievable in the short run, and what was achievable in the longer run. Following DAC guidelines, most evaluations assessed whether project implementation

deadlines were met and budgets were respected, and whether the overall operations' were relevant, efficient and sustainable. However, medium- to long-term impacts were never properly measured for particularly key factors, such as assessing an operation's investment returns. Two facts may explain this omission. First, little *ex ante* economic analysis of the operations was undertaken in most cases, resulting in a lack of both quantifiable definitions of objectives and of baseline data or information from which evaluators could measure impacts. Second, and consequently, there was little economic analysis undertaken in most evaluations, and the use of sophisticated tools, such as economic and econometric modelling, was a rarity. In short, the quantitative and qualitative analyses point to similar conclusions.

Fifth, and finally, the evaluations' conclusions provide little insight as to whether aid for trade works and why. Indeed, evaluators were rarely able to identify causal links between operations and performance. One key problem seems to have been a lack of sustainable financing in many of the operations surveyed, with the result that projects were often terminated prematurely. Another key problem seems to have been the challenge of addressing both economic growth and poverty reduction objectives – when the two can sometimes conflict. Last but not least, there was occasionally a problem of poor project implementation, making it difficult, if not impossible, to assess an operation's impact on trade and poverty-reduction.

However, these findings need to be put into perspective. The failure to refer to specific trade results can be explained at least in part by the absence of closely trade-related objectives in the initial mandate of the operations. Moreover, in the case of 'aid policies' generally - and aid-for trade policies in particular - it is almost always more difficult to assess an operation's sector- or economy-wide impact – than to assess the performance of a specific project in a specific context - given the complex array of extraneous variables influencing outcomes.

It is also possible that the problems outlined above may pertain only to the relatively small sample of evaluations considered in this report. However, the fact that none of the agencies examined appears to perform systematically better than the others suggests that an even broader sample of evaluations would not fare significantly better.

Assuming that the results outlined above are more or less representative of aid-for-trade evaluations generally, we offer the following two broad recommendations: First, every operation should include an *ex ante* assessment of the situation being addressed in order to (i) define objectives in a quantifiable way and (ii) to provide the necessary information (including data) to measure whether these objectives were met *ex post*. Second, evaluators should answer a sequence of questions, set out below, which could provide a framework for assessing operations' impact in a more systematic and thorough way.

## INTRODUCTION

The Aid-for-Trade Initiative has been highly successful in raising the profile of trade as a tool for development. Partner countries have increasingly mainstreamed trade in their development strategies, and donors have responded by mobilising additional resources for trade-related programmes and projects. With public budgets under stress in many developed and developing countries, however, the initiative's continued success requires a better assessment of aid-for-trade outcomes and impacts in terms of trade performances, poverty reduction, growth and development.

As part of the OECD's efforts to improve the effectiveness of aid-for-trade projects and programmes, and to further develop "good practices" in the evaluation of aid for trade, this paper presents a meta-evaluation of some 162 evaluations drawn from the OECD DEReC database. It provides an overview of, and a perspective on, the way that DAC donors and international agencies have implemented operations, and how they have conducted their evaluations - both in terms of methodologies and topics covered.

The paper is organised into four chapters. Chapter 1 explains the purpose, scope and methodology of the study. Chapter 2 attempts to reveal the preferences of evaluators in a quantitative way, based on the frequency with which key words appear in the evaluations. It also attempts to uncover answers to the "what we want to know" question, by focusing on both the trade and the development aspects of aid for trade. Complementing this quantitative approach, Chapter 3 undertakes a qualitative examination of a more narrow set of evaluations that are closely related to aid for trade. Lastly, Chapter 4 tries to answer the "what we would ideally like done" question, by recommending broad guidelines for future evaluations.



## CHAPTER 1. PURPOSES, SCOPE AND METHODOLOGY

### Section 1. Two purposes

The first purpose of this meta-evaluation is to assess whether the selected evaluations offer the information that policy makers in charge of delivering international aid would like to get from field evaluators. In particular, it asks whether trade was an objective of the original operations, and whether trade and development outcomes or impacts were taken into account in the evaluations. To the extent that the latter was the case, it also reports on the kind of trade, poverty reduction and development outcomes described in the evaluations. It should be emphasised that the aim here is not to undertake new evaluations of the operations already evaluated, but rather to examine whether the existing evaluations deliver “what we want to know”.

Ideally “what we want to know” would cover a wide range of issues—from narrower trade outcomes to broader development results. Trade outcomes are important because trade is a proven engine for growth. However, trade is not an end in itself. The ultimate goal of aid-for-trade is to help countries to harness trade behind development and poverty reduction objective – objectives which themselves have many dimensions, from income creation, to income distribution, to broader social concerns (such as gender equality or environmental sustainability).

The second purpose of this meta-evaluation is to propose a set of guidelines (indicators, quantitative and qualitative techniques, etc.) for undertaking future evaluations. In short, to provide a tentative answer to the question “what we would ideally like done”. Two types of guidelines could be proposed that, in our view, are complements rather than substitutes. The first would be general guidelines – involving minimal time, skills and financial resources - to improve the routine evaluation of any operation. The second would be more specific guidelines – involving detailed data collection and specific technical skills – to help undertake more thorough and intensive evaluations of operations' impacts. Since this second approach is much more expensive and time consuming, it is probably applicable only to a carefully chosen subset of all operations. As a result, this paper focuses on the first approach.

### Section 2. Scope

The paper covers a set of evaluations, identified by the OECD Secretariat, covering two countries - Ghana and Vietnam - and two sectors - transport and storage (for the purposes of this analysis, transport and storage have been aggregated into one sector called “TranStor”) - between 1999 and 2010. Transport and storage were selected because these sectors have received significant amounts of aid for trade, and therefore provide a large number of relevant evaluations. Ghana was selected because it is one of the largest recipients of aid for trade in Africa, and has seen its trade expand significantly since 2002. Vietnam was selected because it has served as a model of trade opening for a number of countries, because it has enthusiastically embraced the aid-for-trade agenda, because it has successfully achieved export-led growth, and because it became a WTO member fairly recently, in 2007.

The set of selected evaluations includes only those notified to the DEReC database, not all evaluations undertaken in these countries and in these sectors over the time period. As a crude measure of this study's limited or partial scope, only 162 evaluations are examined in this meta-evaluation,

whereas the total number of operations conducted by DAC members between 1999 and 2008 was 61,677. Hence, the study's results should in no way be interpreted as an exhaustive meta-evaluation of all evaluations of all the operations funded by international aid.

In the current international aid context, operations which attempt to promote economic development by building trade capacity are often referred to as “aid for trade”. In the interests of simplicity, the term aid for trade will be used in this paper to describe all operations that facilitate the integration of developing countries into the global economy through expanded trade (for a more detailed definition, see Chapter 2). However, it should be noted that this term became widely accepted only in the mid-2000s, starting with the discussions surrounding the Millennium Development Goals (see, for instance, Zedillo and Messerlin 2005, Stiglitz and Charlton 2005), in the lead up to the 2005 G8 Summit in Gleneagles, and culminating in the launch of the Aid-for-Trade Initiative at the WTO's 2005 Hong Kong Ministerial Conference.<sup>1</sup>

A key challenge in conducting this meta-evaluation has been to identify the extent to which operations could be analyzed through an aid-for-trade lens. Indeed, some of the evaluations available in the DEReC database for Ghana and Vietnam relate to operations which are likely to impact trade only in the (very) long run – for example, education or health projects. Even in the transport and storage sectors, operations were classified as aid for trade which had at best remote links with international trade. As a result, the study has identified a subset of 43 selected evaluations which have a more direct and immediate bearing on trade (the selection is outlined in Chapter 3). The total stock of evaluations under scrutiny – as well as the narrow sub-set examined in more detail – are more fully described in Annex A.

### **Section 3. Methodology**

A meta evaluation should reveal the implicit interests of evaluators, as well as the “distance” between these interests and the needs of aid-for-trade policy makers. This requires a combination of quantitative measures and qualitative assessments.

The methodology used for quantitative measurement was simple. It relied on the occurrences of words or expressions that could reasonably be considered to be crucial for evaluating aid-for-trade-related operations – and thus that one might expect to see in an evaluation (Annex B lists the 48 words and expressions that were considered key). This method has the advantage of relying on a clearly defined metric: namely the frequency with which key words are mentioned in one hundred pages. Of course, this metric should not be seen as “exhaustive” and “objective”, if only because it is limited to 48 words, and because the words themselves do not necessarily serve as an entirely accurate, or even adequate, measure of the evaluators' interests or policy makers' needs. Yet the word set chosen is probably wide enough to provide meaningful results, and to raise robust questions about the evaluation process in general.

At the same time, this quantitative approach needs to be complemented by a qualitative reading, since the frequency with which words appear in a text may say little about the analytical content of the words themselves, or the context in which they were used. Such a qualitative reading was undertaken for a sub-set of 43 evaluations. The idea was to look beyond the set of issues explicitly treated by the evaluators to see how each of them was being evaluated and to assess the significance of the information that was being provided.

---

<sup>1</sup> As a result, considering the time lag between the implementation of an operation and its evaluation, it is not surprising that operations and evaluations conducted before the mid-2000s do not refer explicitly to the term aid for trade.

## **CHAPTER 2. DO THE EVALUATIONS DELIVER WHAT WE WANT TO KNOW?**

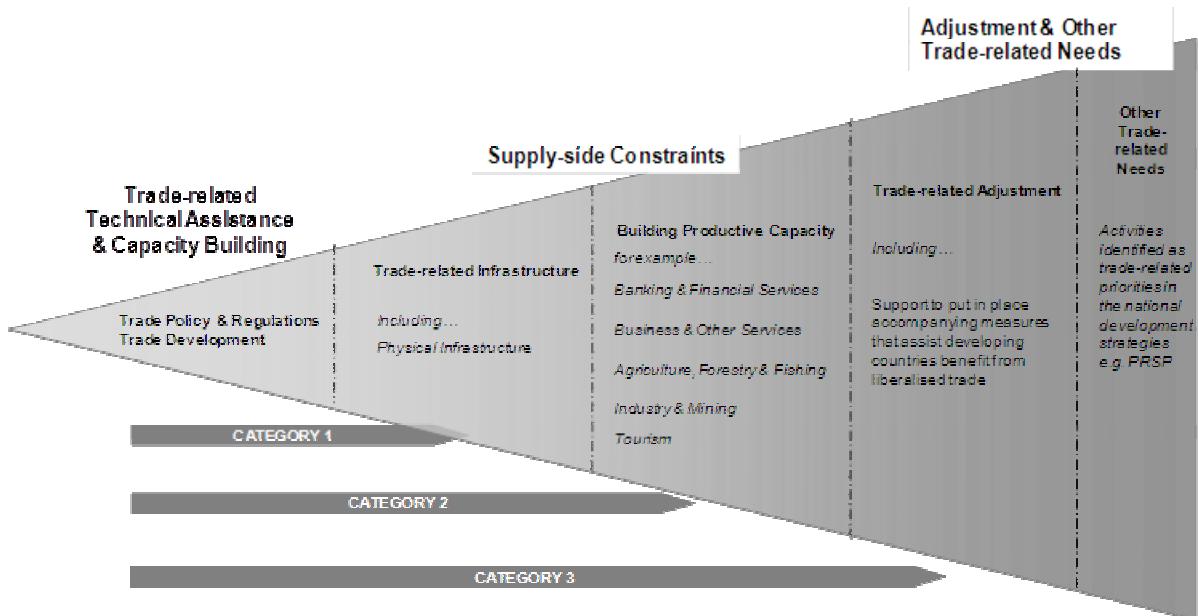
### **Summary of the results**

- Evaluations of aid-for-trade operations did not say much about trade. The words “trade” and “exports” were not among the most frequently mentioned by evaluators, and the WTO or regional trade agreements were largely ignored. More importantly, from an economic point of view, the word “imports” was almost completely ignored.
- Evaluations usually did not clarify the policy linkages which would be of interest to policy makers. For example, the potential impact—positive or negative—of trade policy instruments, such as tariffs, quotas or subsidies, on the return of internationally funded operations was not assessed for the simple reason that these instruments were never examined. Nor were the policy linkages between the operations and “behind the border” policies, such as regulatory reforms or private sector policies, evaluated frequently.
- In sharp contrast with this unexpected silence on trade-related issues, evaluations were much more detailed and diffuse regarding broad development-related concepts. Paradoxically, evaluations tend to rely on general concepts, such as gender or poverty reduction, more often than precise terminology, such as quotas, trade shares, or income distribution, despite near impossibility of using general concepts to describe operations' impacts clearly and robustly.
- The failure to define – or differentiate among – timeframes was widespread. Evaluators rarely distinguished between what was expected in the short run, versus what was expected over the longer run.

### **Introduction**

The chapter adopts a wide definition of the aid-for-trade, following the OECD's [2006] four main categories: trade-related technical assistance (TRTA) and capacity building, trade-related infrastructure, building productive capacity, and macro- and micro-economic adjustment (illustrated in Figure 2.1).

Figure 2.1 Aid for Trade: the agenda



This chapter tries to give a sense of whether the selected evaluations deliver information relevant to the aid-for-trade agenda. Section 1 focuses on the trade dimension, while section 2 focuses on the development dimension. Both sections assess how frequently key words and expressions occur in the evaluations—a crude measure of the revealed interests of the evaluators – and try to determine whether or not regular patterns emerge. For instance, does a donor use the same set of words in its evaluations regardless of country or sector – signalling the use of precise guidelines in evaluations? Or do the donor's interests appear to shift according to the country or sector under consideration – suggesting the absence of firm guidelines?

In order to determine whether an underlying pattern emerges, an arbitrary convention is used. It is assumed that 25 occurrences or more of a given word or expression per 100 pages represents a substantial and permanent interest in the word under scrutiny - and hence of the issue that it is supposed to represent. Such occurrences are coloured in the following tables in order to get a “graphic” view of the bigger picture.

### Section 1. The trade component

If one wants to define what policy makers reading evaluations would like to learn about the trade aspects of aid-funded operations, four main topics should be raised:

- i) the extent to which evaluations capture the role of trade;
- ii) the extent to which they take into account the trade policy of the country in question;
- iii) the extent to which they refer to the trade-related issues specific to aid for trade (such as technical barriers to trade, sanitary and phytosanitary standards, TRTA, adjustment policies, trade facilitation, etc.); and

- iv) the extent to which they take into account key international trade agreements.

The analysis of these four questions relies on the frequency of key words and expressions (the number of times a word has been used per one hundred pages of evaluations). It is first analysed on a country/sector basis, and then on a donor basis. It is also analyzed for the overall set of 162 evaluations, as well as for the narrower, more “aid for trade-related” set of evaluations. As mentioned above, the qualitative analysis in Chapter 3 focuses on a narrow set of 43 evaluations drawn from the larger set.<sup>2</sup>

### ***To what extent do evaluations capture the role of trade?***

This broad question can be split into more precise questions, each captured by a key word or expression. Have evaluators simply mentioned the word “trade”? Or have they been more precise by referring to the two sides of trade—“exports” and/or “imports”? The words exports and imports themselves have very different connotations. One would expect exports to be a visible trade indicator in aid-funded operations aimed at a better integrating developing countries into the world economy – and indeed the aid for trade agenda typically focuses on the importance of exports. However, from an economic perspective, imports are ever more important than exports because they are the channel through which cheaper and/or better goods, services, investments and innovations are made available to domestic consumers— whether households or firms. In other words, imports are a critical input for future growth and development, while exports are mostly the means by which countries buy much needed imports.

### ***The recipient perspective***

Table 2.1A presents the frequency with which the key words attached to the above questions appeared in the evaluations, aggregated by recipients. Three main observations arise: First, broadly speaking, the trade aspect - defined as the combined occurrences of the words “trade”, “exports” and “imports” - was relatively prominent (frequencies above 50 occurrences per 100 pages are rarely observed for other key words or expressions, as shown below). Moreover, the frequency with which trade terms were mentioned was substantially higher for Vietnam than for Ghana.

Second, the focus was clearly on exports rather than imports, reflecting a possible lack of economic understanding of the latter's role in growth and development.

Third, the trade and export terms appeared relatively more frequently in the narrow set of evaluations than in the overall set for the two countries and sectors. This seems to justify the decision to look beyond the overall set of evaluations, and to focus on a narrower set for the qualitative assessment (see Chapter 4). A similar pattern does not emerge across the TranStor sector. Because many TranStor operations relate mainly to domestic, as opposed to international, infrastructure (*i.e.*, roads or railways might link two cities but have little direct impact on trade), one might have expected even bigger differences between the two sets of evaluations, but this does not appear to have been the case.

---

<sup>2</sup> There is a minor difference between the narrow set used in this chapter and the one used in Chapter 4. The evaluations done by IOB (Ghana) and EBRD, NEI and NORAD (Vietnam) are not reported in this chapter because these four donors have reported to DEReC only one evaluation for the three fields (Ghana, Vietnam and TranStor).

**Table 2.1A. The role of trade by recipient**

	Overall set			Narrow set		
	Ghana	Vietnam	Transtor	Ghana	Vietnam	Transtor
How trade is mentioned?						
Is "trade" mentioned?	8.2	31.0	22.2	13.0	159.6	43.6
Which component of trade is mentioned: exports?	15.7	12.8	7.9	27.7	60.9	12.7
Which component of trade is mentioned: imports?	6.2	5.6	3.9	10.1	12.2	6.5
How is trade understood?						
as a macro-economic issue: trade balance?	0.1	0.0	0.0	0.1	0.0	0.0
as a pure trade issue: comparative advantages?	1.8	1.5	0.7	1.7	1.1	1.7
as a production issue: diversification?	1.8	0.7	0.1	2.8	1.0	1.0
as a welfare issue: gains from trade?	0.0	0.0	0.0	0.0	0.0	0.0

Source: DEReC database. Authors' computations

Beyond observations about how frequently trade was mentioned in the evaluations, it is also interesting to note how evaluators perceived the relationship between trade and other policies. This question has four dimensions: (i) Do evaluators see trade as a macroeconomic issue (an approach which would be revealed by the frequency of reference to trade balances)? (ii) Do they see trade as critical to helping countries strengthen comparative advantages in the world economy? (iii) Do they see trade as a means of diversifying domestic production? (iv) And do they see trade leading to a reallocation of domestic resources in a way that increases a country's overall welfare? The answers to these four questions reveal the implicit "policy linkages" that evaluators might have in mind.

The second part of Table 2.1A suggests that evaluators were not much concerned with clarifying such policy linkages - if indeed these linkages were even considered. Specifically, it was not clear that trade was perceived as a macroeconomic issue – or that macroeconomic imbalances were seen as a trade policy concern. This neglect is somewhat worrisome, particularly for country-case evaluations.

### ***The donor perspective***

Table 2.1B presents the frequency with which key words appeared in evaluations of donor operations. The overall set of evaluations rendered several interesting insights. First, the same patterns observed in Table 2.1A were repeated for almost every donor, suggesting a "parallel" behaviour amongst donors when evaluating their operations in a given country or sector. In other words, the observations drawn from Table 2.1A were not artificial averages. Such parallel behaviour is easy to explain, since the questions related to how often trade was mentioned. What is more puzzling was evaluators' low level of interest in clarifying trade's implicit policy linkages.

**Table 2.1B. The role of trade by donor, the overall set of evaluations**

	Overall set of evaluations										Narrow set of evaluations									
	Ghana										Ghana									
	AIDB	AFD	CIDA	DANIDA	DFID	EC	Japan	SIDA	USAID	WB	All	AIDB	AFD	DANIDA	DFID	EC	WB	All	W. std	
How trade is mentioned?																				
Is "trade" taken into consideration?	2.2	10.3	2.6	6.3	6.7	54.0	0.0	1.9	1.5	2.6	6.7	1.8	2.1	6.9	4.7	9.1	54.0	25.0	11.1	1.2
Which component of trade is mentioned: exports?	17.7	19.0	0.0	7.0	3.0	58.9	0.0	0.0	0.5	6.1	8.7	1.6	29.6	19.0	8.3	4.7	58.9	84.1	17.9	0.9
Which component of trade is mentioned: imports?	27.3	12.1	28.9	1.7	40.6	4.0	26.9	27.9	24.9	28.8	23.9	0.6	5.8	0.0	2.0	2.0	4.0	50.0	4.2	1.8
How is trade understood?																				
as a macro-economic issue: trade balance?	0.2	0.0	0.0	0.0	0.1	1.7	0.0	0.0	0.0	0.0	0.1	2.6	0.0	0.0	0.0	0.2	0.8	0.0	0.2	1.9
as a pure trade issue: comparative advantages?	4.6	0.0	0.0	1.5	1.9	2.2	0.0	2.9	0.3	1.4	2.1	1.0	6.9	0.0	0.4	2.0	2.4	11.4	2.3	1.2
as a production issue: diversification?	2.2	8.9	5.3	1.0	0.4	24.2	0.0	0.0	0.0	0.3	2.0	1.8	2.6	8.6	1.0	0.5	24.2	4.5	3.7	1.3
as a welfare issue: gains from trade?	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	--	0.0	0.0	0.0	0.0	0.0	0.0	--	
Vietnam																				
	AIDB	AFD	CIDA	DFID	EC	GTZ	Japan	SECO	SIDA	WB	J-Eval	All	W. std	AIDB	EC	GTZ	SECO	SIDA	All	W. std
How trade is mentioned?																				
Is "trade" taken into consideration?	41.9	7.5	2.9	15.7	193.7	30.8	0.0	199.2	9.6	17.0	8.1	31.0	1.6	167.1	340.4	35.3	199.2	14.1	159.6	0.9
Which component of trade is mentioned: exports?	11.2	16.3	0.0	7.2	19.3	12.5	0.0	246.2	4.7	7.2	1.2	12.8	2.4	31.4	36.4	13.0	246.2	2.6	60.9	1.5
Which component of trade is mentioned: imports?	5.9	0.0	5.9	0.0	12.1	0.0	0.0	5.4	6.4	4.6	0.2	5.6	1.1	20.4	11.1	0.0	5.4	1.3	12.2	1.1
How is trade understood?																				
as a macro-economic issue: trade balance?	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	0.0	0.0	0.0	0.0	0.0	0.0	--
as a pure trade issue: comparative advantages?	0.6	1.3	0.0	1.2	2.9	0.0	0.0	3.8	2.1	0.7	0.7	1.5	1.0	0.8	0.0	0.0	3.8	1.3	1.1	1.3
as a production issue: diversification?	1.8	0.0	0.0	0.0	0.0	1.7	0.0	0.0	0.3	0.0	0.0	0.7	2.1	1.5	0.0	1.7	0.0	0.0	1.0	1.4
as a welfare issue: gains from trade?	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	--	0.0	0.0	0.0	0.0	0.0	0.0	--	
Transtor																				
	AsDB	AIDB	DANIDA	GTZ	Japan	SIDA	WB				All	W. std	AsDB	AIDB	DANIDA	Japan	WB	All	W. std	
How trade is mentioned?																				
Is "trade" taken into consideration?	30.5	26.8	6.9	0.0	0.0	2.8	16.2				22.2	1.1	51.3	29.4	27.7	0.0	31.1		43.6	0.7
Which component of trade is mentioned: exports?	7.9	13.8	2.0	0.0	26.3	0.7	8.5				7.9	1.1	34.1	14.4	8.5	26.3	6.7		12.7	0.7
Which component of trade is mentioned: imports?	5.0	8.0	0.5	0.0	5.3	0.0	1.6				3.9	1.1	7.6	10.0	2.1	5.3	1.1		6.5	0.7
How is trade understood?																				
as a macro-economic issue: trade balance?	0.0	0.2	0.0	0.0	0.0	0.0	0.0				0.0	2.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	--
as a pure trade issue: comparative advantages?	0.8	1.0	1.0	0.0	0.0	0.0	0.2				0.7	1.1	1.8	3.1	4.3	0.0	0.4	1.7	0.9	
as a production issue: diversification?	0.8	1.3	1.0	16.7	0.0	0.0	0.0				1.0	2.2	1.4	1.3	0.0	0.0	0.0	1.0	1.4	
as a welfare issue: gains from trade?	0.0	0.0	0.0	0.0	0.0	0.0	0.0				0.0	--	0.0	0.0	0.0	0.0	0.0	0.0	0.0	--

**Notes:** WB: World Bank, J-Eval: joint evaluations. W.Std: weighted standard deviations

Source: DEReC database. Authors' computations.

Two additional patterns are worth noting. First, there was a wider spectrum of frequencies for Vietnam than for Ghana (as revealed by higher variances for Vietnam on exports). At first glance, this might seem to be the result of the dominance of one donor, SECO, in Vietnam. However, eliminating SECO does not dramatically narrow this difference. Second, the three joint evaluations in Vietnam were in the low band of frequencies.

The narrow set of evaluations provides interesting additional information. It amplifies massively the reference to "trade" and "exports", except for Ghana. But there was almost no change regarding how trade – and its implicit policy linkages – was understood by evaluators. All these observations will be substantiated in more detail in the qualitative analysis in Chapter 4.

### ***To which extent do evaluations refer to the trade policy of countries?***

Another relevant question is whether evaluators understood a country's broader trade and economic policy context, especially in the TranStor sector, and whether they appreciated its relationship to – and impact on – operations. For example, the returns from aid-funded operations might vary greatly depending on whether a country imposed moderate and stable tariffs, or high and volatile tariffs. Another important variable is whether the country is landlocked. Reducing tariffs could provide only small benefits if transport costs are high and transport infrastructure is poor or non-existent. Such constraints can represent "tax-equivalents" many times higher than import tariffs. By the same token, even sizable infrastructure investments can generate low returns if existing high tariffs are not reduced.

In this context, two successive questions should be asked. Have evaluators simply looked at “trade restrictions” without examining their nature and context in more detail? Or have they gone further and looked in detail at the three main trade policy instruments—tariffs, quotas and subsidies)—and their potentially very different effects and costs?

### ***The recipient perspective***

Turning to evaluations on the recipients’ side, the results presented in Table 2.2A give rise to two observations. First, evaluators largely ignored the general term “trade restrictions”, despite the potential impact such restrictions could have on the operations’ returns. This was the case for both the broad and the narrow sets of evaluations. The fact that this observation can be made for the narrow set of evaluations is especially significant, given that a key criterion in selecting the narrow set in the TranStor sector (as explained in Chapter 3) was an explicit international trade dimension in the evaluations.

**Table 2.2A. The role of trade policy, by recipient**

	Overall set			Narrow set		
	Ghana	Vietnam	Transtor	Ghana	Vietnam	Transtor
In broad terms: trade restrictions?	0.1	0.0	0.0	0.1	0.0	0.0
In terms of specific instruments?						
tariffs?	2.4	8.2	9.9	3.3	12.0	12.8
quotas?	0.2	0.5	0.1	0.2	0.1	0.0
subsidies?	2.2	3.4	3.1	2.7	2.8	3.7

Source: DEReC database. Authors’ computations

Second, evaluators were only marginally more aware of the existence of trade instruments. Tariffs showed a low frequency, and subsidies were even lower, despite the fact that most aid-funded operations in the TranStor sector ultimately have a substantial subsidy component. Turning to subsidies, evaluators tended to limit their analysis to a comparison of subsidy schemes provided by various countries. They stop short of assessing whether such subsidies were desirable or not—their indirect message being that subsidies were desirable only if they did not endanger the government budget (see illustrations for Ghana and Vietnam in Chapter 3).

Third, evaluators were largely silent on quotas, although this instrument is likely to have the most perverse and negative development impact, since quotas are opaque, can generate high distortions, and encourage costly rent-seeking behaviour among domestic and foreign firms alike.

Lastly, the awareness of such instruments was higher in the case of Vietnam - a country that was in the process of undertaking much faster and deeper liberalization than Ghana. This may reflect the fact that Vietnam’s trade policy was rapidly evolving during the period, whereas Ghana’s trade policy was less active. At the same time, it raises serious questions about the usefulness of evaluations if they do address existing policies that can have a direct bearing on the success of the operations.

### ***The donor perspective***

Table 2.2B shifts the focus to donors – and asks whether the above observations apply uniformly to all donors or whether there are “atypical” cases.

Once again donors show parallel behaviour, as illustrated by low variances. In particular, no evaluator paid attention to trade restrictions in the TranStor sector for the two countries. Variances were higher for tariffs and subsidies, with AfDB and AsDB acknowledging the importance of tariffs in Ghana and Vietnam, and only the AFD highlighting the relevance of subsidies (with the above *caveat* on the impact analysis). The TranStor case was not comparable because the word “tariff” was often used as an equivalent of the word “price” (e.g., “water tariff”).

**Table 2.2B. The role of trade policy, by donor**

**Notes:** WB: World Bank, J-Eval: joint evaluations, W.Std: weighted standard deviations.

Source: DEReC database. Authors' computations

### *To what extent do evaluations refer to trade-related issues specific to aid for trade?*

As illustrated in Figure 2.1, the aid-for-trade agenda covers a much wider set of trade-related policies than the basic trade instruments examined in the previous paragraphs. These issues include: (i) "technical barriers to trade" (TBT) for industrial goods, (ii) "sanitary and photo-sanitary standards" (SPS) for agricultural products, (iii) trade facilitation, (iv) adjustment policies (especially those associated with helping countries adjust to trade opening), and (v) trade assistance. This section examines whether these issues emerged in the selected evaluations.

### *The recipient perspective*

Table 2.3A suggests that, within the overall set of evaluations, these issues were, by and large, ignored. However, within the narrow set of evaluations, references to SPS standards, trade facilitation (in the TranStor sector), adjustment policies (in Ghana), and trade assistance (in Vietnam) surface more frequently.

**Table 2.3A. The “Aid for Trade” wider set of trade issues, by recipient**

	Overall set			Narrow set		
	Ghana	Vietnam	Transtor	Ghana	Vietnam	Transtor
Technical barriers to trade?	0.0	1.6	0.0	0.0	11.4	0.0
Sanitary and phytosanitary standards?	0.0	2.6	0.0	0.0	18.3	0.0
Trade facilitation?	0.0	3.3	3.9	0.1	23.5	11.3
Adjustment policies?	0.0	3.6	4.0	12.9	2.7	4.3
Trade assistance?	0.0	1.8	0.0	0.0	12.5	0.1
Trade-related technical assistance?	0.0	0.2	0.0	0.0	1.6	0.0

Source: DEReC database. Authors' computations

Some of these results are puzzling. The almost total absence of references to TBTs, SPS standards, and trade-related technical assistance is particularly surprising, given the high priority that both recipients and donors typically place on these topics. It should be noted that the relative frequency with which trade facilitation was mentioned in Vietnam may be less impressive than it appears, given that Vietnam was the second-largest recipient of trade facilitation-related aid in the world between 1999 and 2008, with 65 projects (or 2.6 per cent of all projects in the world Source: Database CRS/DAC OCDE). Similarly, it is surprising that adjustment policies were mentioned more frequently for Ghana than for Vietnam by a significant margin, given that Ghana had a more stable product composition of exports and imports than Vietnam over the decade examined, particularly in the most recent years. This heightened sensitivity to adjustment issues in Ghana may reflect the long and bitter debate in many African countries about the gains from trade and donors' "adjustment programmes" of the 1980s and 1990s (see the donor perspective below).

### ***The donor perspective***

What are the similarities and dissimilarities among donors? Limited variance (if any) could be expected among donors for topics showing low frequencies. However, the narrow set of evaluations revealed some substantial differences among the donors. The high score of adjustment policies in the World Bank's evaluations of its Ghana operations may partly reflected sensitivity towards the above mentioned debate on adjustment policies in Africa. Similarly, the very high score for TBT and SPS in the EC's and SECO's evaluations of their Vietnam operations may reveal Europe's obsession with these issues.

**Table 2.3B. The “Aid for Trade” wider set of trade issues, by donor**

	Overall set of evaluations								Narrow set of evaluations															
	Ghana				Vietnam				Transtor				Ghana											
	AsDB	AIDB	AFD	CIDA	DANIDA	DFID	EC	Japan	SECO	SIDA	USAID	WB	Joint Eval	All	W. std	AsDB	AIDB	AFD	DANIDA	DFID	EC	WB	All	W. std
Technical barriers to trade?	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	
Sanitary and phytosanitary standards?	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	
Trade facilitation	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	3.2	0.0	0.0	0.0	0.8	0.0	0.1	2.4		
Adjustment policies?	3.7	0.0	2.6	5.8	1.7	7.3	0.0	1.9	0.0	9.8	4.7	1.0	3.7	0.0	6.5	1.7	7.3	45.5	5.7	1.6				
Trade assistance?	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	0.0	0.0	0.0	0.8	0.0	0.0	0.0	-	
Trade-related technical assistance?	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	
Technical barriers to trade?	0.0	0.0	0.0	0.0	41.5	0.0	0.0	5.4	0.0	0.0	0.0	0.0	1.6	2.9		0.0	86.9	0.0	5.4	0.0	0.0	11.4	2.1	
Sanitary and phytosanitary standards?	0.0	0.0	0.0	0.0	37.7	0.0	0.0	54.6	0.0	0.0	0.0	0.0	2.6	2.3		0.0	78.8	0.0	54.6	0.0	0.0	18.3	1.4	
Trade facilitation	10.4	0.0	0.0	0.0	1.0	0.0	7.7	2.3	0.0	0.0	0.0	0.0	3.3	1.9		47.4	2.0	0.0	2.3	0.0	0.0	23.5	2.0	
Adjustment policies?	2.9	1.3	5.9	1.2	1.0	3.3	0.0	0.8	1.9	7.2	18.9	3.6	1.3	4.1	1.0	3.5	0.8	0.0	0.0	0.0	2.7	1.0		
Trade assistance?	0.0	0.0	0.0	0.0	48.8	0.0	0.0	0.8	0.0	0.0	0.0	0.0	1.8	3.3		0.0	102.0	0.0	0.8	0.0	0.0	12.5	2.2	
Trade-related technical assistance?	0.0	0.0	0.0	0.0	5.3	0.0	0.0	1.5	0.0	0.0	0.0	0.0	0.2	2.6		0.0	11.1	0.0	0.0	0.0	0.0	1.6	2.2	

**Notes:** WB: World Bank, J-Eval: joint evaluations, W.Std: weighted standard deviations.

Source: DEREc database. Authors' computations

### *Do evaluations take into account key international trade agreements?*

The “aid-for-trade” agenda has emerged largely independently from the Doha Round and negotiations of regional trade agreements. Nevertheless, the debate has permeated most trade negotiations over the last decade. Thus, it is useful to look at whether references to key trade agreements were mentioned in the evaluations. The obvious candidate is the WTO’s Doha Round of multilateral negotiations which have largely continued throughout the period covered (starting with the 2001 Doha Ministerial Conference and continuing passed the 2008 Geneva mini-Ministerial Meeting). In addition to the WTO, references to regional or preferential trade agreements would also be significant, since such agreements have echoed, and often amplified, WTO discussions on aid for trade.

#### *The recipient perspective*

Table 2.4A reveals an almost complete lack of reference to key trade agreements, with the exception of the WTO; and even the WTO references are limited to operations in Vietnam. This observation is not surprising, given that Vietnam’s on-going WTO accession negotiation was a key policy issue throughout the decade covered.<sup>3</sup> The absence of references to trade agreements in Ghana’s evaluations may reflect Ghana’s resistance to accepting new commitments during the Doha

<sup>3</sup> Vietnam’s negotiations for accession to the WTO started in January 1995 and were completed in November 2006. Vietnam is undergoing a long (yet unfinished) process of implementing its commitments. For instance, as shown in Table A1 (Annex A), Vietnam’s final level of bound tariffs will be 11.4 percent, while the 2008 level was still 16.8 percent.

negotiations. It may also reflect Ghana's relatively limited involvement in regional or bilateral agreements, and the modest impact of those on Ghana's trade policy (for instance, ECOWAS has had a very limited effect on Ghana's already moderate tariffs, while the Cotonou Convention imposed no discipline on Ghana's trade policy during the decade covered). In the case of the TranStor sector, the absence of references to trade agreements underlines the basic difficulty of negotiating services liberalization, particularly transport services (which have long been ruled by special international schemes geared towards limiting international competition), in most trade agreements.

**Table 2.4A. International trade agreements, by recipient**

	Overall set			Narrow set		
	Ghana	Vietnam	Transtor	Ghana	Vietnam	Transtor
WTO?	0.2	6.9	0.5	0.3	41.5	1.5
Regional trade agreements?	0.0	0.1	0.0	0.0	0.4	0.0
Preferential trade agreements?	0.0	0.0	0.0	0.0	0.4	0.0

Source: DEReC database. Authors' computations

### **The donor perspective**

Regarding trade agreement references in evaluations of donor operations, one finds a low frequency in both the overall and narrow sets. The one notable exception was the high frequency with which the WTO was mentioned in EC and SECO evaluations of Vietnam operations. This was in sharp contrast to the low profile of these key words in the Ghana evaluations.

**Table 2.4B. International trade agreements, by donor**

	Overall set of evaluations									Narrow set of evaluations											
	Ghana									Ghana											
	AfDB	AFD	CIDA	DANIDA	DFID	EC	Japan	SIDA	USAID	WB	All	W. std		AfDB	AFD	DANIDA	DFID	EC	WB	All	W. std
WTO?	0.0	0.0	0.0	0.0	0.1	4.0	0.0	0.0	0.0	0.0	0.2	3.0		0.0	0.0	0.0	0.2	4.0	0.0	0.5	2.3
Regional trade agreements?	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-		0.0	0.0	0.0	0.0	0.0	0.0	0.0	-
Preferential trade agreements?	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-		0.0	0.0	0.0	0.0	0.0	0.0	0.0	-
	Vietnam									Vietnam											
	AsDB	AfDB	CIDA	DFID	EU	GTZ	Japan	SECO	SIDA	WB	All	W. std		AsDB	EC	GTZ	SECO	SIDA	All	W. std	
WTO?	0.3	0.0	0.0	13.3	137.7	0.8	0.0	41.5	1.5	0.0	0.7	6.9	2.3	1.3	278.8	0.9	41.5	2.6	41.5	1.9	
Regional trade agreements?	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.1	3.3		0.0	3.0	0.0	0.0	0.0	0.4	2.2	
Preferential trade agreements?	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		0.3	0.0	0.0	0.0	0.0	0.1	2.2	
	Transtor									Transtor											
	AsDB	AfDB	DANIDA	GTZ	Japan	SIDA	WB			All	W. std			AsDB	AfDB	DANIDA	Japan	WB	All	W. std	
WTO?	1.0	0.2	0.0	0.0	0.0	0.0	0.0			0.5	2.2			2.3	0.0	0.0	0.0	0.0	1.5	2.2	
Regional trade agreements?	0.0	0.0	0.0	0.0	0.0	0.0	0.0			0.0	-			0.0	0.0	0.0	0.0	0.0	0.0	-	
Preferential trade agreements?	0.0	0.0	0.0	0.0	0.0	0.0	0.0			0.0	-			0.0	0.0	0.0	0.0	0.0	0.0	-	

**Notes:** WB: World Bank, J-Eval: joint evaluations, W.Std: weighted standard deviations.

Source: DEReC database. Authors' computations.

## **Section 2. The development component**

This section shifts the focus from trade-related issues to those more closely related to development. Three questions seem particularly relevant in this context:

- do evaluators take into account the broad constraints that an inefficient public sector could impose on development?

- ii) do evaluators take into account the broad constraints related to inefficiencies in the private sector?
- iii) and do evaluators pay enough attention to the broad performance of the economy or sector examined?

***How much attention is paid to public sector-related constraints?***

An inefficient public sector can impose significant constraints on growth and development – and these constraints can be even more difficult to surmount in developing countries than in developed countries. From a development perspective, it is thus essential that evaluators give some sense of their perception of whether public sector-generated constraints exist, and whether such constraints had a substantial impact on the returns of the operations evaluated. Against this background, the following section looks at another spectrum of policy linkages—between aid-funded operations, infrastructure policies and regulatory policies.

Two questions are relevant when examining public sector-generated constraints. First, do evaluators focus on the consistencies – or inconsistencies – among physical infrastructure projects that might impact on the success of the operations evaluated? Second, do evaluators focus on “regulatory infrastructure” issues that also impact on the success of operation? To operate efficiently, firms need appropriate domestic regulations as well as appropriate roads or bridges and this “regulatory capital” can play as important a role as physical or human capital in economic development. The following section divides regulatory policies into three subsets: (i) the specific regulations closely related to the evaluated operations; (ii) the broad regulatory framework of the country or sector; and (iii) the even broader issue of “governance”.

***The recipient perspective***

Table 2.5A provides an interesting picture of how these various public sector-related constraints are understood by evaluators, with both the broad and the narrow sets of evaluations giving similar results (with the usual very limited magnification effect in the narrow set). First, it is clear that evaluators are concerned with physical infrastructure policies. The systematically higher frequency with which physical infrastructure was mentioned in the TranStor sector compared to the country cases suggests that evaluators are even more inclined to take a broad view of the inter-action between an operation and other physical infrastructure when they are already involved in sectoral infrastructure policy. However, the lower frequency for country cases may simply reflect evaluators' inability to take a “bigger picture” view of economies operating with many interactive agents (*i.e.*, a “general equilibrium” perspective).

**Table 2.5A. Public sector-related constraints, by recipient**

	Overall set			Narrow set		
	Ghana	Vietnam	Transtor	Ghana	Vietnam	Transtor
Does the evaluation mention physical infrastructure problems?	15.3	21.7	38.0	20.5	46.2	66.5
Does the evaluation refer to specific regulations?	4.7	9.1	9.0	7.2	12.8	9.5
Does the evaluation refer to the broad regulatory framework?	0.9	9.7	1.9	1.5	1.7	2.4
Does the evaluation raise broad governance issues?	20.4	20.3	5.0	18.9	6.8	8.5

Source: DEReC database. Authors' computations.

Second, in sharp contrast, evaluators clearly showed less interest in regulatory issues—whether specific or general. For Ghana in particular, references to specific regulations were infrequent, while references to the broad regulatory framework was almost non-existent.

Third, references to broad governance issues were more frequent across evaluations. However, the problem, as underscored in the qualitative analysis, is that evaluators, with a few notable exceptions, often said little in detail about governance. The general impression is of evaluators paying lip service to a topic that is acknowledged to be crucial for economic growth and development, but on which they had little concrete or constructive to say.

Lastly, it is worth noting an interesting pattern that emerged from the data. Regulatory issues (specific and general) had notably more salience in Vietnam than in Ghana. This may reflect the impression that Vietnam's regulatory policies are somewhat better defined than Ghana's policies. By contrast, broad governance issues seemed to carry more weight in Ghana than in Vietnam. This may reflect the broad government's inability to design appropriate regulations.

### ***The donor perspective***

Given that the words and expressions related to public sector-constraints appear fairly frequently, one might expect substantial differences among donors.

The overall set of evaluations revealed that only one word – *i.e.*, infrastructure for TranStor – reached the threshold of 25 occurrences per 100 pages across most donor evaluations. The rest of the picture is patchy, with no donor expressing the same interest in a word over the evaluations for the two countries and the TranStor sector. If there was a pattern, it was that regional development banks focussed on infrastructure in the countries of their region (*i.e.*, AfDB for Ghana, AsDB for Vietnam). Another pattern was that governance issues tended to attract the most attention from different donors – with CIDA and SIDA focusing on governance in Ghana, and DFID focusing on governance in Vietnam. Finally, those evaluators who show the least interest in governance issues (AFD, GTZ and Japan) were among those who showed the most interest in infrastructure.

**Table 2.5B. Public sector-related constraints, by donor**

	Overall set of evaluations										Narrow set of evaluations					Ghana						
	Ghana					Ghana					Ghana					Ghana						
	A S D B	A F D B	C I D A	D A N I D A	D F D	J a p a n	S I D A	U S A I D	W B	A L L	W . S t d	A S D B	A F D B	D A N I D A	D F D	E C	W B	A L L	W . S t d			
Mention of physical infrastructure problems?	38.8	0.0	7.9	20.5	3.0	37.1	7.7	2.9	2.5	14.2	14.9	1.1	45.5	0.0	24.2	4.2	37.1	43.2	21.9	0.8		
Does the evaluation refer to specific regulations?	5.9	0.0	2.6	13.0	3.6	6.5	0.0	0.0	0.0	3.1	5.4	1.2	6.9	0.0	15.4	5.2	6.5	13.6	9.5	0.7		
Does the evaluation refer to the broad regulatory framework?	0.6	0.0	0.0	0.3	0.4	16.9	0.0	0.0	0.0	0.2	1.0	2.9	1.1	0.0	0.4	0.5	16.9	0.0	2.0	2.1		
Does the evaluation raise broad governance issues?	9.3	0.0	123.7	25.2	29.0	33.1	0.0	54.8	35.3	12.1	24.1	1.1	12.7	0.0	25.4	31.4	33.1	59.1	26.2	0.7		
Vietnam																						
	A S D B	A F D B	C I D A	D A N I D A	D F D	E U	G T Z	J a p a n	S E C O	S I D A	W B	J-E v a l	A L L	W . S t d	A S D B	A F D B	D A N I D A	D F D	E C	W B	A L L	W . S t d
Mention of physical infrastructure problems?	50.1	7.5	2.9	14.5	7.2	1.7	23.1	13.1	8.3	37.3	3.9	21.7	1.0	90.6	0.0	17	13.1	2.6	46.2	1.8		
Does the evaluation refer to specific regulations?	6.0	0.0	5.9	72.3	17.9	0.8	0.0	1.5	5.2	29.4	29.1	9.1	1.4	11.0	12.1	32.2	6.1	3.8	12.8	0.9		
Does the evaluation refer to the broad regulatory framework?	13.9	0.0	2.9	1.2	17.4	30.8	0.0	6.2	6.9	2.0	8.1	9.7	1.2	2.8	2.0	0.9	0.0	0.0	1.7	1.1		
Does the evaluation raise broad governance issues?	23.0	23.8	29.4	4.8	29.5	115.8	46.2	20.0	13.3	35.3	17.6	20.3	0.9	8.2	3.0	0.9	13.8	1.3	6.8	1.0		
Transtor																						
	A S D B	A F D B	D A N I D A	G T Z	J a p a n	S I D A	W B	A L L	W . S t d	A S D B	A F D B	D A N I D A	J a p a n	W B	A L L	W . S t d						
Mention of physical infrastructure problems?	32.9	49.7	33.5	100.0	94.7	13.8	72.1			38.0	0.6	55.3	40.6	93.6	94.7	114.1			66.5	0.4		
Does the evaluation refer to specific regulations?	8.1	4.1	8.9	16.7	0.0	6.2	16.6			9.0	0.7	10.1	2.5	12.8	0.0	11.9			9.5	0.8		
Does the evaluation refer to the broad regulatory framework?	1.1	0.2	2.5	0.0	0.0	0.0	8.8			1.9	1.8	1.4	0.6	10.6	0.0	5.6			2.4	1.2		
Does the evaluation raise broad governance issues?	5.0	5.4	3.4	0.0	0.0	1.7	7.9			5.0	0.9	8.0	0.6	14.9	0.0	14.4			8.5	0.9		

**Notes:** WB: World Bank, J-Eval: joint evaluations, W.Std: weighted standard deviations.

Source: DEReC database. Authors' computations.

The narrow set of evaluations provided a more nuanced – and interesting – picture. Generally speaking, the narrow set amplified indications that donors showed significant interest in infrastructure, especially with the emergence of the World Bank as a key player. It also magnifies the apparent consensus among donors that governance is a key issue in Ghana.

### **How much attention is paid to supply-side constraints?**

The economic literature on development gives great weight to the importance of supply-side constraints in the economic growth equation. Since public sector-related constraints are only one aspect of supply-side obstacles, the following section focuses on private sector-related constraints, while acknowledging that in practice the line between public and private supply-side constraints is often difficult to draw clearly.

Since the aid-for-trade literature repeatedly underscores the importance of supply side constraints, one might expect evaluators to have focussed extensively and systematically on this issue. One might also expect evaluators to have provided information on expenditures or costs—either the direct costs involved in completing the operation, or, more interestingly from the perspective of this section, the indirect costs to the whole economy of not (or if only partially) completing an operation. In this context, one might also expect evaluators to give some sense of whether the private sector grew as a result of the operation. Finally, one might also ask whether evaluators showed a systematic interest in technical assistance in the broader sense of the term (*i.e.*, not just trade-related technical assistance).

### ***The recipient perspective***

Table 2.6A provides a mixed picture. First, it shows that evaluators failed to address the issue of supply-side constraints, both in the overall and the narrow set of evaluations.

Second, it shows that evaluators focused on expenditures unevenly. Only in the case of Ghana were expenditures mentioned frequently, and then only by four donors (Danida, DFID, the EC and the World Bank). Moreover, evaluators used the term expenditures in two different ways: first, in the context of expenditures exceeding budgets; and second, to differentiate between an operation's recurring (often assumed to be covered by the recipient) and non-recurring expenditures (often funded by the donor).<sup>4</sup>

Third, although the term “private sector growth” appeared rarely in both the overall and the narrow sets of evaluations, the term “private sector” appeared much more frequently, reflecting the fact that it was often used in neutral statements (*e.g.*, encouraging the development of the private sector by enhancing the business environment in which it operates) and not in the context of evaluating specific operations (see Chapter 3).

**Table 2.6A. Supply-side constraints, by recipient**

	Overall set			Narrow set		
	Ghana	Vietnam	Transtor	Ghana	Vietnam	Transtor
Does the evaluation mention supply-side constraints?	0.0	0.0	0.0	0.0	0.0	0.0
Does the evaluation refer to expenditures (costs)?	34.2	10.8	14.7	37.9	2.7	13.8
Does the evaluation refer to private sector growth?	0.2	0.0	0.1	0.3	0.2	0.1
Does the evaluation deal with technical assistance?	15.5	13.6	26.0	20.4	21.3	28.8

Source: DEReC database. Authors' computations.

### ***The donor perspective***

Shifting to donors, Table 2.6B shows that for the broad set of evaluations, none of the keywords reached the threshold of 25 occurrences per 100 pages, both for the two countries and for the TranStor sector. The terms “supply-side constraints” and “private sector growth” were uniformly absent in all evaluations. The term “expenditures” was mentioned only in the case of Ghana. The rest of the picture is fragmented.

<sup>4</sup> It is worth mentioning that the term was used to refer to public expenditures. It is particularly true for the TranStor sector for which several evaluations stressed the lack of involvement of the private sector in the management of infrastructure.

**Table 2.6B. Supply-side constraints, by donor**

	Overall set of evaluations								Narrow set of evaluations							
	Ghana				Vietnam				Transtor				Ghana			
	AID B	AID F	CIDA	DANIDA	DFID	EU	Japan	SECO	SIDA	USAID	WB	ALL	WB	ALL	WB	ALL
Does the evaluation mention supply-side constraints?	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	-	-	-
Does the evaluation refer to expenditures (costs)?	10.9	8.6	10.5	37.0	56.2	26.6	7.7	0.0	2.5	51.5	36.4	1.0	12.2	8.6	43.3	26.6
Does the evaluation refer to private sector growth?	0.3	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.3	0.1	1.5	0.5	0.0	0.2	0.0
Does the evaluation mention technical assistance?	6.5	13.8	13.2	26.7	14.2	23.4	0.0	0.0	5.0	13.0	14.8	0.8	6.9	13.8	31.3	14.5
	Ghana								Vietnam							
	AID B	AID F	CIDA	DANIDA	DFID	EU	Japan	SECO	SIDA	USAID	WB	ALL	WB	ALL	WB	ALL
Does the evaluation mention supply-side constraints?	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Does the evaluation refer to expenditures (costs)?	6.1	6.3	5.9	12.0	3.9	2.5	0.0	8.5	6.5	62.7	46.0	10.8	1.4	1.8	0.0	2.6
Does the evaluation refer to private sector growth?	64.2	1.3	0.0	0.0	1.4	0.8	0.0	3.1	0.7	0.0	0.2	20.5	2.9	0.5	0.0	0.0
Does the evaluation mention technical assistance?	27.0	18.8	0.0	6.0	25.6	2.5	0.0	11.5	4.0	13.1	19.2	13.6	0.9	32.7	26.3	2.6
	Vietnam								Transtor							
	AID B	AID F	CIDA	DANIDA	DFID	EU	Japan	SECO	SIDA	USAID	WB	ALL	WB	ALL	WB	ALL
Does the evaluation mention supply-side constraints?	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	0.0	0.0	-
Does the evaluation refer to expenditures (costs)?	14.2	7.4	20.7	0.0	0.0	7.2	11.2	0.0	0.0	0.0	0.0	14.7	0.9	14.6	5.6	17.0
Does the evaluation refer to private sector growth?	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	2.6	0.2	0.0	0.0
Does the evaluation mention technical assistance?	31.8	11.1	13.8	0.0	0.0	24.1	21.6	0.0	0.0	0.0	0.0	26.0	0.8	38.0	9.4	8.5
	Transtor								Ghana							
	AID B	AID F	CIDA	DANIDA	DFID	EU	Japan	SECO	SIDA	USAID	WB	ALL	WB	ALL	WB	ALL

**Notes:** WB: World Bank, J-Eval: joint evaluations. W.Std: weighted standard deviations.

Source: DEReC database. Authors' computations.

A similar pattern emerges for the narrow set of evaluations, with the exception of the expenditure item for Ghana. All of the words continued to have very low frequencies in the evaluations, even though, as noted above, they reflect key priorities in the aid-for-trade agenda.

### **How much attention is paid to performance?**

The basic aim of evaluations is to provide a sense of whether operations - projects, programmes or general aid policies – performed according to expectations. Given that development is such a broad subject, it more difficult to come up with a concise or exhaustive list of words and terms that might reflect evaluators concerns with development outcomes.

However, the following words – ranging from more precise to more general – were identified as key to such a list: economic growth, competitiveness, efficiency, effectiveness, sustainability, poverty reduction and gender gap. A brief description of these terms – and the rationale behind their selection – is provided in Annex B.

### **The recipient perspective**

The frequency with which key development-related words appeared in evaluations of donor operations is presented in Table 2.7A. One clear observation emerged: namely, that the more general words and expressions were mentioned much more frequently than the more precise ones. On the surface, this observation is puzzling, given that it is presumably harder for evaluators to assess an operation using general words and expressions, such as poverty reduction or gender, rather than by using more precise ones, such as economic growth. The answer may lie in the fact that, as illustrated

by the qualitative analysis (see Chapter 3), evaluators typically used these words and expressions to state an operation's objectives, not to assess an operation's performance in meeting these objectives.

**Table 2.7A. Development performances, by recipient**

	Overall set			Narrow set		
	Ghana	Vietnam	Transtor	Ghana	Vietnam	Transtor
Economic growth?	6.0	5.8	7.9	8.1	12.4	12.4
Competitiveness?	0.3	2.4	1.6	0.5	12.0	2.5
Efficiency?	11.4	11.2	24.0	9.7	18.7	34.5
Effectiveness?	24.8	19.0	16.3	24.5	22.6	19.1
Sustainability?	24.0	15.0	26.8	27.1	24.4	33.4
Poverty reduction?	23.9	21.5	14.4	27.7	8.1	24.6
Gender gap?	19.2	20.2	8.3	22.1	6.5	3.9

Source: DEReC database. Authors' computations.

### ***The donor perspective***

The frequency with which key development-related words appeared in evaluations of recipient operations is presented in Table 2.7B. Several observations can be drawn from it. First, the terms economic growth and competitiveness were rarely mentioned by evaluators, with one exception, GTZ in Vietnam. Second, the other key words and expressions were used more frequently, but generally did not reach the threshold of 25 occurrences per 100 pages. Third, the pattern was somewhat different for Ghana, on the one hand, than for Vietnam and the TranStor sector, on the other. In the Ghana's case, two donors, CIDA and DANIDA, used the widest range of key words most frequently. In the case of Vietnam and the TranStor sector, however, the picture was more fragmented, with several donors using a wide range of words, but fairly infrequently (e.g., the EC for Vietnam).

The narrow set of evaluations amplifies the observations made for the overall set. Most of the donors focussed on sustainability, poverty reduction and the gender gap in the case of Ghana. But in the case of Vietnam and the TranStor sector, the frequencies were scattered across donors.

**Table 2.7B. Development performances, by donor**

	Overall set of evaluations										Narrow set of evaluations									
	Ghana					Vietnam					Ghana									
	AfDB	AFD	CIDA	DANIDA	DFID	EC	Japan	SIDA	USAID	WB	ALL	W. std	AfDB	AFD	DANIDA	DFID	EC	WB	ALL	W. std
Refer to economic growth?	17.7	5.2	5.3	6.0	4.2	12.1	3.8	0.0	1.0	2.1	5.7	0.9	17.5	5.2	6.7	4.9	12.1	6.8	8.1	0.6
Refer to competitiveness?	0.0	0.0	0.0	0.7	0.1	3.2	0.0	0.0	0.0	0.3	0.4	2.3	0.0	0.0	0.8	0.2	3.2	0.0	0.7	1.8
Refer to efficiency?	37.0	0.0	26.3	12.7	15.8	0.0	0.0	0.0	0.0	10.3	13.6	1.3	33.9	0.0	7.9	18.9	0.0	0.0	13.6	1.4
Refer to effectiveness?	19.9	0.0	36.8	24.0	42.5	27.4	15.4	3.8	7.0	24.5	26.0	0.7	19.6	0.0	22.8	40.3	27.4	11.4	26.8	0.7
Refer to sustainability?	49.4	19.0	50.0	29.2	17.1	45.2	11.5	20.2	15.4	10.6	23.7	0.6	38.1	19.0	26.2	18.7	45.2	45.5	27.7	0.4
Refer to poverty reduction?	64.9	0.0	68.4	32.5	21.0	89.5	26.9	1.9	0.5	12.6	27.8	1.0	65.6	0.0	30.7	21.6	89.5	38.6	37.3	0.8
Refer to gender gap?	30.7	0.0	44.7	34.2	17.5	38.7	3.8	26.9	7.0	7.6	20.8	0.7	25.4	0.0	38.8	12.8	38.7	25.0	26.8	0.6
Transtor																				
	AsDB	AfDB	DANIDA	GTZ	Japan	SECO	SIDA	WB	J-Eval	ALL	W. std	AsDB	AfDB	DANIDA	SECO	SIDA	WB	ALL	W. std	
Refer to economic growth?	10.6	3.8	0.0	4.8	9.2	1.7	7.7	2.3	3.1	11.1	3.0	5.8	0.7	15.6	16.2	1.7	2.3	34.4	12.4	1.0
Refer to competitiveness?	2.0	0.0	0.0	0.0	6.8	44.2	0.0	6.2	1.0	0.0	0.5	2.4	2.4	6.9	8.1	46.1	6.2	2.6	12.0	1.3
Refer to efficiency?	18.4	1.3	14.7	8.4	27.5	4.2	53.8	10.8	4.5	9.8	19.4	11.2	0.9	22.7	41.4	43	10.8	3.8	18.7	1.0
Refer to effectiveness?	20.8	7.5	35.3	95.2	44.9	5.0	69.2	17.7	12.7	13.1	30.0	19.0	0.9	24.2	49.5	52	17.7	14.1	22.6	0.8
Refer to sustainability?	18.0	15.0	35.3	4.8	37.7	30.0	36.5	28.5	9.5	19.0	16.4	15.0	0.5	13.0	50.5	31.3	28.5	32.1	24.4	0.4
Refer to poverty reduction?	25.2	0.0	5.9	59.0	32.9	2.5	0.0	0.8	10.4	102.6	51.5	21.5	1.2	9.2	12.1	2.6	0.8	17.9	8.1	0.8
Refer to gender gap?	5.3	1.3	108.8	6.0	18.8	0.8	0.0	2.3	28.0	3.3	50.8	20.2	1.6	0.3	26.3	0.9	2.3	28.2	6.5	1.2

**Notes:** WB: World Bank, J-Eval: joint evaluations. W.Std: weighted standard deviations

Source: DEReC database. Authors' computations.

### Section 3. Concluding remarks

Does this quantitative exercise offer any common lessons? Or - perhaps more accurately, given the limited results from both the overall and narrow sets of evaluations - does it offer any salient “questions”? The following two lessons/questions emerge as the most important ones.

First, there were few clear links between a donor and specific issue, meaning there were few systematic “leaders” in terms of evaluations on a given topic, or for a given country and/or sector. One donor might pay attention to a particular issue for one country, but then completely ignore that same issue for another country (or sector). Systematic evaluations should thoroughly examine all potential issues and topics; and if one key topics is ignored, it should be for clearly stated reasons. If this interpretation of the data is correct, then it argues strongly for a serious effort to establish more detailed guidelines - and a stricter system of reviews - for evaluations (which in turn might allow greater freedom and creativity in evaluations).

Second, the key words and expressions that are most precise, from an economic analysis perspective, tended to appear less frequently. Moreover, evaluations gave little information on the broader economic context in which the operations were taking place (the “general equilibrium” perspective), making it difficult to clarify key policy linkages.

This was particularly striking for the linkages that would have been easiest to assess. For instance, the fact that evaluators ignored trade policy instruments suggests that they had little understanding of – or gave little weight to – the extent to which the rate of return of operations could be reduced or increased by such instruments. High tariffs on inputs crucial to an operation would obviously limit its effectiveness, just as high tariffs imposed on foreign substitutes to an operation's

output could artificially increase the operation's benefits. Including this kind of analysis in evaluations would be useful for recipients in particular, allowing them see just how damaging high tariffs are to a country's interests. It would also be useful for donors, underlining the contradictions and costs of inconsistent policies—*e.g.*, advocating freer trade in general, while accepting a high tariff structure in a particular case if it boosts the return on donor operations.

## CHAPTER 3: QUALITATIVE ANALYSIS

### Summary of the results

- For all the issues considered, the qualitative analysis in this chapter largely echoes the conclusions reached in the quantitative analysis in the previous chapter.
- The qualitative analysis reveals that an operation's impact on trade was clearly not the focus of evaluators' work. Indeed, in a number of cases, the trade impact was never explicitly addressed.
- While some terms indicating an attention to policy linkages appeared in a number of evaluations, the qualitative analysis reveals that these terms were not well defined and their use was not substantiated by reference to hard evidence.
- The qualitative analysis reveals that evaluations focused more on broad, development-related concepts than on precise trade issues – and that even the development-related concepts were interpreted vaguely and narrowly. For instance, gender was either mentioned in passing as a cross-cutting issue, or measured in terms of short-term and irrelevant objectives that had little meaningful bearing on women's economic or societal situation (such as the number of women who attended seminars, as part of an operation).
- The qualitative analysis reveals that evaluators rarely incorporated relevant time dimensions in their assessments, and suggests two key reasons for this: First, little *ex ante* economic analysis of the operations was undertaken, meaning that the operation's objectives remained un-quantified and there was no baseline information from which evaluators could measure an operation's impact. Second, and to a large extent consequently, there was little economic analysis undertaken in most evaluations.

### Introduction

A “quantitative” analysis of key words and expressions works best in a large set of evaluations. This chapter presents a “qualitative” analysis of the narrow set of 43 selected evaluations (an evaluation on Vietnam for 2010 was added in the qualitative analysis because of some particularly interesting features, see below). Its main aim is to test the observations made in the previous chapter.

With the goal of applying a coherent methodology across country-related and sector-related evaluations, the following series of 49 questions was applied to the narrow set of evaluations to explore the characteristics of each operation, and to assess the evaluations' outcomes. The narrow set of 43 selected evaluations was not based on word frequencies but on readings insights.

<b>Characteristics of the evaluated projects</b>	<i>Type of evaluation (next)</i>
1 Aid for Trade focused? 2 Trade-related? 3 Product-specific? 4 Which products? 5 Project, programme or general aid policy? 6 Population specific? 7 Which population? 8 Donor specific? 9 Country specific? 10 Date of implementation of the project under scrutiny?	26 Population specific? 27 What population? 28 What product? 29 Is the timeliness evaluated? 30 Is the respect of the budget evaluated? 31 Qualitative? 32 Field mission? 33 Quantitative? 34 Primary data collection? 35 Indicators? 36 Econometric analysis? 37 Estimation of benefits? 38 Is there any type of robustness analysis? 39 Is the impact on the environment estimated? 40 Is the socio-economic impact estimated? 41 Is the impact on governance and institutions estimated? 42 Is the rate of return calculated?
<b>Type of evaluation</b>	
11 Is the project's success assessed? 12 Is the project definition/conception evaluated? 13 Is the impact in trade explicitly considered? 14 Is the impact on trade explicitly measured? 15 Is the impact on poverty reduction explicitly considered? 16 Is the impact on poverty reduction explicitly measured? 17 Period under scrutiny? 18 Part of a repeated evaluation process or one shot? 19 Do they relate to previous evaluations? 20 Is the initial situation well identified? 21 Is there any type of benchmarking? 22 Is causality identified in any way? 23 Explicated methodology? 25 Product-specific?	43 Is the project found to be successful? 44 Is the impact on trade found to be positive? 45 Is the impact on poverty reduction found to be positive? 46 Rate of return? 47 Any delay? 48 Length of the delay? 49 Budget respected?
<b>Evaluation outcome</b>	

However, this common framework needed to be tailored to the specific countries and sectors, and when applied to a particular set of evaluations. The framework also needed to be adjusted to fit the specific design and purpose of certain studies.

A DEReC evaluation can examine either a specific project or a broader “review” of a collection of projects (grouped according to specific sectors, such as infrastructure or private sector development, or according to themes, such as trade facilitation or poverty reduction). Reviews usually provide a long-term perspective,<sup>5</sup> and they rely, not just on evaluations, but also on desk studies. One example is AsDB’s review of its infrastructure development operations under the umbrella of the “Greater Mekong Sub-region programme”. Focusing on the theme of trade facilitation, the review considered

---

<sup>5</sup> Thus, reviews are different than programme evaluations. One such example is the review of ADB’s infrastructure development and the “Greater Mekong Subregion programme”. Through the lenses of trade facilitation the review considers several projects implemented between 1990 and 2009 and refers to various thematic impact assessments, for example the impact on the reduction of time travel but also on the spread of HIV and the impact on poverty reduction. Such ADB “special evaluation studies” assess the impact of all ADB operations in one sector or under one type of programme. SIDA uses the same type of approaches when assessing their Business Training programme considering operations from 1998 to 2006.

several infrastructure projects implemented between 1990 and 2009, and encompassed various thematic impact assessments, such as travel times, the spread of HIV, or poverty reduction.

The framework was adjusted depending on the purpose of the project and whether it had a direct aid for trade-orientation. For each country and sector, the following sections will explain the characteristics of the selected evaluations, the methodology followed, and the issues covered. Each section then assesses the overall aid-for-trade impact of the operations.

## Section 1. Aid for Trade in Vietnam

### *The set of selected evaluations*

Out of the 64 evaluations made available by the OECD (specific to Vietnam or to a set of countries including Vietnam), ten were selected and classified according to the WTO Aid for Trade Task Force's definition of aid for trade [OECD, 2006]. Thus these 10 evaluations were grouped into three categories: trade-related technical assistance (TRTA), infrastructure and private-sector development. The international dimension was taken as the first discriminator. Operations focused on the domestic market were not included in the selection, even though they could have some spill-over effects on trade (e.g. a project aimed at producing paper for the Vietnamese market).

Table 3.1 presents these ten evaluations. Trade clearly did not figure nearly as prominently in evaluations of private sector development projects compared to evaluations of TRTA and infrastructure projects. The qualitative analysis of Vietnam's narrow-set of evaluations reveals that only "reviews" looked at trade enhancement and poverty reduction in any detail. This was mainly because reviews are concerned with long-term impact assessments, whereas the focus of project evaluations tends to be more short term. It is also worth noting that such reviews are specific to AsDB.

**Table 3.1. Key data for the narrow set of evaluations: Vietnam**

Evaluation Category	Evaluation date	Institution	Period under scrutiny	Programme/ Project	Review/ Evaluation	Word occurrences: frequency per 100 pages				
						trade	export	import	Sustain-ability	Poverty
TRTA	2007	SECO-UNIDO	2004-2006	Project	Evaluation	35	88	2	16	6
	2008	EU	2005-2008	Project	Evaluation	340	36	11	51	31
	2009	SECO	2007	Project	Evaluation	119	33	14	5	5
	2009	SECO	2004-2007	Project	Evaluation	372	462	5	48	5
Infrastructure	2003	ADB	1990-2000	Project	Evaluation	24	35	28	33	2
	2008	ADB	1992-2008	Programme	Review	299	49	30	13	21
	2010	ADB	1992-2009	Programme	Review	246	90	33	0	315
Private Sector Development	2006	SIDA	1998-2004	Programme	Evaluation	14	3	1	32	32
	2006	ADB	1983-2006	Programme	Review	24	6	4	6	18
	2009	GTZ	2005-2009	Programme	Evaluation	32	13	0	31	4
Average all evaluations)						35	13	3		

Another important point worth noting is that all the selected evaluations scored a positive rating, even if some included less satisfactory points. That said, there were notable differences among the various categories of operations evaluated. Some based their assessments just on field missions (consisting in interviews with non-statistically selected stakeholders), whereas others also made use of objective indicators, such as the rates of return in infrastructure and private sector development operations. The extent to which donors harmonise their operations with each other and with the government was also usually noted - and positively rated - in the evaluations.

#### ***Assessment of the Aid for Trade dimension: methodologies and issues covered***

The methodologies used by evaluators were highly dependent on the type of operation being assessed. For example, quantitative tools were barely used for assessing TRTA operations, but were used extensively for evaluating infrastructure operations. However, two general comments apply to all selected evaluations.

First, an operation's impact on producers, exporters, and importers was hardly ever evaluated, with the exception of a few case-studies in the SECO-UNIDO evaluation. According to many evaluators, this reflected the failure to collect relevant data during the operations' implementation, and the insufficient time between the end of operations and the start of evaluations, especially for TRTA evaluations. Another reason for this omission was that the direct beneficiaries of most operations was governments - as well as a few private-public partner organizations – so evaluations tended to be short-term and focussed on government clients.

Second, few studies made use of economic models to assess the impact of infrastructure programmes on poverty reduction; the one study that did was not an evaluation *stricto sensu*, but rather a prospective analysis. In none of the other evaluations were econometric tools employed to identify causality. Indeed, the AsDB and SIDA frequently made a point of mentioning the need for more robust analysis in their evaluations. Only the “reviews” addressed trade or socio-economic impacts – and here the impact assessment was usually subjective.

#### ***TRTA operations***

Four TRTA operations were implemented to strengthen institutional reform in Vietnam, as part of the country's effort of open its economy and accede to the WTO. Each was focused to setting up workshops or developing certification and testing facilities – all with the overarching objective of sharing information on the implications of WTO rules, and helping Vietnam to move up the export value chain. There was a clear bias in this set of TRTA evaluations, in that three of the four were sponsored by SECO either as the principal donor or as a partner with UNIDO or SIDA. The fourth evaluation assessed an EU funded operation.

All of these evaluations were carried out by independent consultants. They followed donors' guidelines and analyzed the operations' impact, relevance, efficiency and sustainability, as well as their effect on a series of broader, cross cutting issues, such as gender or environment.

All the evaluations were qualitative, based mainly on field missions. Evaluators frequently complained about being unable to assess the outcome of operations more thoroughly because the failure to collect baseline data against which progress could have been measured through indicators. This suggests that project conception should be clarified from the start, with the basic information and data needed to assess a project's success provided in advance. The sustainability of operation outcomes was one of the most poorly documented dimensions. Not only was it never supported by quantitative evidence, but the concept itself was poorly understood – with the focus on the sustainability of the

project (*i.e.*, whether budgets were renewed or operational authority was successfully transferred to Vietnamese officials or partner organizations) rather than on the sustainability of outcomes (notably in terms of trade).

Last but not least, the link between an operation's implementation and poverty reduction was considered only in the context of the project objectives - and the results were never assessed. The causal link between the various TRTA projects and trade development was never clearly established, other than by relying on the evaluator's intuition.

The operations' outcomes were assessed according to two criteria: first, the effective implementation of the project, such as the number of participants at workshops (a poor indicator of participants' capacity to implement any knowledge gained); and, second, interviews with stakeholders to determine their relative satisfaction. Outcomes were never assessed in terms of the effective transfer of knowledge or the operations' direct impact on trade. In fact, assessments were largely subjective, with evaluators often repeating the expression "in our view".

### ***Infrastructure operations***

All three evaluations referenced in Table 3.1, under the infrastructure category, were AsDB evaluations. Two of them relate to the same regional programme. One of the three was a working paper by independent authors employed by the AsDB Institute, while the other two were undertaken by the AsDB's Operations and Evaluation Department. It is not surprising that the last two evaluations follow a similar pattern, based on a well-defined methodology, and systematically rate the relevance, effectiveness, efficiency, impact and sustainability of the projects' objectives and outcomes. Moreover, the evaluations look at comprehensive programmes—not stand alone projects—meaning they should be classified as long term "reviews", with evaluators assessing the impact of the whole endeavour rather than the impact of a single and short-term initiative.

Generally these evaluations employ a more robust impact assessment methodology. When assessing the impact of infrastructure projects, AsDB evaluators collected primary data from implemented projects. They undertook quantitative data collection, such as traffic or border-crossing point counts, as well as qualitative surveys (called "perception surveys"), with the aim of assessing the impact of trade facilitation projects on the movement of goods and people across borders. In contrast, much of the information used for assessing the impact of operations on socioeconomic parameters, gender, and environment was secondary data gathered from government and private-sector officials. Unfortunately, such survey data does not lend itself to thorough econometric analysis. However, it does allow evaluators to estimate some economic benefits that will flow from a project's implementation. A good illustration is the evaluation of Saigon's port rehabilitation project. Evaluators were able to forecast cargo handling performances - and the evolution of various other flows – based on whether or not the project was implemented - a task made possible because key indicators were already defined and available pre-implementation as part of the initial project outline. However, the evaluation team was not able to identify the extent to which these economic benefits were passed on to the Vietnamese economy as a whole.

Three factors could explain why evaluators use quantitative tools more extensively in evaluations of infrastructure projects than in evaluations of TRTA operations. First, the time frame for infrastructure projects is much longer than for TRTAs projects (the two AsDB infrastructure projects covered more than ten years), meaning that evaluators have scope to employ more sophisticated assessment tools. Second, donors are keen to ensure the profitability of their investments in large-scale infrastructure projects, and they typically insist on the importance of gathering *ex ante* information - and employing relatively rigorous methodologies - when evaluating the impacts of the projects they

financed. Finally, the outcomes of such operations (*e.g.*, whether the number of cargo ships using new harbour facilities increased) are easier to monitor and evaluate quantitatively than the outcomes of TRTA operations (*e.g.*, whether entrepreneurs and or officials usefully employed the knowledge gained in workshops).

The AsDB review of the “Greater Mekong Subregion” (2008) deserves special mention. It was one of the most thorough evaluations covered by this meta-evaluation. Not only was the time frame long enough for evaluators to measure impacts quantifiably, but the review also included various impact studies directly relevant to the programme. Although these impact studies had more than one focus (*i.e.*, socio-economic, trade, investment or poverty reduction issues), they provided evaluators with valuable information based on robust methodologies which, when taken together, offered a comprehensive assessment of the programme’s overall impact. For instance, the AsDB (2010) document quantified the links between infrastructure investment and poverty reduction, using a multi-region, general equilibrium model, supplemented with household survey data (it did not assess the effective impact on poverty but rather estimated its impact on an *ex ante* basis).

As for the trade impact, evaluators consistently failed to make the link between infrastructure initiatives and programmes aimed at developing trade and investment. For example, the evaluators of the Greater Mekong Subregion programme assessed the success of infrastructure development and the creation of transport corridors, but they pointed to the need to transform this transport corridor into an “economic corridor.” However, the available case studies analyzing the impact of the project could already establish the link between the operation and the development of trade within each country. New projects focusing on trade facilitation would, in their view, enlarge this impact to inter-country trade.

As shown in the analysis of the frequency with which the word “trade” is mentioned in the evaluations, there is a tendency for evaluators to focus more on “exports” than on “imports”. In this respect, the AsDB infrastructure reviews and evaluations stand out as exceptions. The evaluation of Saigon’s port rehabilitation, for example, highlighted the importance of imports for Vietnamese businesses, and pointed out that a decrease in import prices of key inputs, such as such as fertilizer, steel or iron, was crucial for the development of Vietnam’s economy. Moreover, the same observations made for Vietnam were also made for the TranStor evaluations.

### ***Private sector development operations***

Three evaluations relating to operations on private sector development – undertaken by SIDA, AsDB and GTZ - were selected from the pool of all the available evaluations. The qualitative analysis of these three private-sector development evaluations offers the same basic impression as the earlier analysis of word occurrences. In none of the evaluations, with the possible exception of the GTZ study, was trade explicitly considered, nor was the impact on poverty ever assessed, even in the longer-term reviews.

The AsDB evaluation is a wide review of all its private-sector development operations from 1985 to 2006. Thus Vietnam is just one case study among several, and the assessment of project performance relies mostly on an examination of investment performance. The focus is on performance and strategy, not impacts. Only a few details are mentioned about one AsDB project that, according to the Operation Evaluation Department, created jobs and had an overall positive socio-economic impact.

The SIDA and GTZ evaluations suffer from the same shortcomings as the TRTA operations’ evaluations. That is to say, they rely mostly on qualitative information derived from field evaluations and interviews with beneficiaries, none of which were selected on the basis of objective statistical

methods. Following donor guidelines, evaluators made an effort to address the projects' broad impact, in areas such as poverty reduction, the environment, gender, sustainability and relevance. However, the lack of necessary data, combined with the insufficient time between the projects' implementation and the evaluations, precluded any quantitative measurement of impacts. As a result, for example, the gender impact (when mentioned) was assessed according to the number of female trainees participating in workshops. SIDA evaluators frankly complained about the lack of control groups or primary data collection which would have allowed them to assess the direct impact of the projects – especially as Vietnam was experiencing a period strong economic growth at the time. That said, the two evaluations contained a more robust quantitative analysis than the TRTA evaluations, and a more comprehensive attempt to assess longer-term issues. For example, the SIDA evaluators mentioned, not only the direct beneficiaries of the private-sector development operation, but also the “ultimate” intended beneficiaries, meaning small- and medium-sized enterprises (SMEs) (although they were unable to assess the impacts). Evaluators also noted the importance of allowing sufficient “maturity time” for entrepreneurs to absorb and benefit from programmes before attempting to assess their impacts – and the impossibility of drawing any useful conclusions when time frames were too short. As a result, the evaluators felt unable to infer that SMEs growth was a consequence of any specific project outcome. The same weakness hampered the GTZ evaluator's capacity to assess the impact of programmes aimed at employment and private-sector development.

### ***Outcomes of evaluations: does aid for trade work?***

Reaching clear conclusions about what works and why in aid for trade remains difficult. Existing evaluations tend to tell us much more about how an operation was implemented - and its economic sustainability (with information on rates of return, for example) - than about how it impacted on trade or poverty reduction. From this narrower perspective, however, most selected evaluations were positively rated.

It is important to highlight once again the uniqueness of Vietnam in this meta-evaluation, whether in terms of the clear linkages between Vietnam's WTO accession and the implementation of the aid-for-trade agenda, or the Government's firm commitment to trade and business development. In short, Vietnam served as an ideal “laboratory” for the aid-for-trade agenda.

Despite the positive synergies in Vietnam, however, evaluators proved unable to provide a robust impact assessment of aid for trade on trade performance, and even less on poverty reduction. Even though the relevance of many operations was beyond doubt in this context, it remained difficult to infer any direct causality between the implementation of aid-for-trade operations and Vietnam's positive economic and trade dynamic.

### ***Final remarks***

The above analysis highlights the importance of distinguishing between evaluations and reviews (that is, between short-term and long-term evaluations). Long-term studies were much more relevant in terms of analyzing the extent of one donor's action and its impact on cross-cutting issues.

Cross-cutting issues (e.g., the environment, HIV/AIDS, gender, etc.) are mandatory points that evaluators must consider in order to follow donors' guidelines and fulfil their terms of reference. However, most of the time, evaluators highlight the irrelevance of such issues to project outcomes, or the lack of data and information needed to undertake assessments. Because of its specific setting, the Greater Mekong programmes evaluation was able to address the impact of improved infrastructure on cross-cutting issues, such as HIV/AIDS. When HIV/AIDS was mentioned in other evaluations, however, it was considered irrelevant to the type of operations evaluated. As for the environment,

apart from the AsDB evaluation of Saigon's port rehabilitation project, this issue was only mentioned as a consequence of economic development in general. The important link between environmental issues, economic development and infrastructure improvements was never disputed in the evaluations, but the impact assessment was usually too abstract, too general, and dependant on too many other factors to be relevant. As for gender issues, they were almost always mentioned in evaluations, but through the lens of irrelevant or unusable statistics, such as the percentage of women participating in various training workshops.

## **Section 2. Aid for Trade in Ghana**

### ***The set of selected evaluations***

Out of the 41 evaluations available for Ghana in the DeReC database, seven had to be eliminated either because they were not *ex post* evaluations (but rather *ex ante* examinations of operations to decide whether they should be financed), or because they were not concerned directly with Ghana.

Of the remaining 34 evaluations, 13 were selected for the qualitative analysis, spread equally across projects, programmes, and aid policies. Given the small number of evaluations available, and their tenuous links to trade, the selection procedure was based on a simple criterion. As long as an operation was aimed at facilitating the integration of developing countries into the global economy through the expansion of trade (even if this objective was marginal), it was selected. As a result, among the 13 evaluations selected, only a fraction of them can be said to evaluate an aid-for-trade operation *stricto sensu*. Had a stricter criterion been used, the number of selected evaluations would have been too small to draw statistically meaningful conclusions. What is more, even the most trade-related evaluations chosen cannot really be considered aid for trade-focused. For example, none of the evaluations has a true international trade dimension; when international trade issues are considered, it is only “*en passant*”.

It should be noted that this broad definition of aid for trade used in selecting the narrow set of evaluations for Ghana is very different from the definition adopted for Vietnam and TranStor evaluations. As a result, the data can be compared only with caution. To some extent, the lower frequency of “trade words” in the Ghana set is to be expected since the evaluations were not directly concerned with aid for trade.

As shown in Table 2.2, the 13 operations evaluated were strongly biased towards aid policies, which account for over half of the selection, and represent two thirds of the total “aid policy” evaluations available in the stock. This unsurprisingly shows that most aid policies tend to have some aid for trade component. This also shows that less frequent trade-related word occurrences are to be expected because evaluations of this type are usually longer and concerned with a wider range of issues. A third of operations evaluated in the selection were programmes (this also represents a third of such evaluations available in the total stock), while the remaining two-thirds were concerned with specific projects (accounting for only a sixth of the available project evaluations in the stock). To the extent that the evaluations in the DeREC database are representative of operations being undertaken and evaluated in Ghana, it seems clear that donors do not focus on aid-for-trade projects in Ghana. This might be due to the poverty-reduction priorities of the Government and some donors and/or to a wide reluctance to recognise the potentials gains from trade. However, it is difficult to reach any firm conclusions given that there is no complete list of the operations which donors have undertaken and evaluated, so the representativeness of the DeREC database cannot be assessed.

## ***Assessment of the aid-for-trade dimension: methodologies and issues covered***

### ***Issues covered: the elusive quest for trade***

The issues that the evaluations covered closely followed the guidelines adopted by donors (which in turn reflect DAC guidelines). As a result, evaluations most often investigated the impact of operations on poverty reduction, the environment, social and gender issues, as well as on private sector participation – as in the case of Vietnam.<sup>6</sup>

In contrast, the trade impact was not always explicitly considered. In fact, although all 13 evaluated operations dealt with trade-related issues, such as supply-side constraints or transport and energy market issues, none were explicitly designed to advance trade objectives. What is more, when the impact on trade was mentioned, it was most often couched in theoretical or hypothetical terms, and never systematically measured or verified. Returning to the word count exercise, the total occurrence of the word trade within evaluations selected is 13 times per 100 pages on average, half the figure for poverty reduction. This gives a good idea of the evaluators' limited focus on trade.

In addition, the qualitative reading of the 13 selected evaluations indicates that the word trade does not occur in trade-impact analysis, but rather in the country context provided by evaluators. In the “aid policy” evaluations reviewed, only one had a paragraph explicitly evaluating the impact of several operations on international trade (those related to transport). It also contained a few paragraphs examining the impact of operations on private-sector development, and on the diversification of Ghana’s productive/export base. However, these paragraphs were relatively short, and, like the other selected evaluations, did not contain quantitative impact measurements.

### ***Methodology used in the narrow set of evaluations***

As mentioned above, most of the evaluations were based on a set of criteria that closely followed DAC guidelines. However, these criteria were used: (i) to look at short-term impacts only, or at the characteristics of the operations; and (ii) in the absence of techniques that would have allowed for the evaluation of performance in a satisfactory and comparable way. Regarding the first issue, most evaluations assessed the relevance of operations in term of whether they fit the country context, addressed the Government’s objectives and strategies (*i.e.*, the policy linkages), and met budget, disbursement and implementation targets. Regarding the second issue, the evaluations’ impact findings were often hypothetical or superficial, and lacked objective measurement. Indeed, none of the 13 evaluations used quantitative tools, and the evaluation processes never involved data collection, construction of indicators, or quantitative techniques. Lack of data was frequently mentioned as a shortcoming of the evaluations—especially the lack of baseline data that would have made possible a comparison of the *ex ante* and *ex post* situations. When evaluators described their methodologies, they relied exclusively on one or both of the following tools: (i) the analysis of documents obtained from officials and sector stakeholders, and (ii) the analysis of the results of field interviews with relevant stakeholders and/or focus groups. These same tools were also being used informally in evaluations where no particular methodology was identified.

Such practices are problematic in two respects. First, it is difficult, if not impossible, to assess operations’ outcomes through non-quantitative methods alone. Second, the non-quantitative analysis is itself challenging and problematic, as the evaluators themselves acknowledged. Not only were key documents often unavailable, but evaluators rarely explained how those documents that were available

---

<sup>6</sup> In AfDB evaluations, community participation and regional integration are also considered – the latter being of interest in an Aid for Trade perspective (although, as will be shown below, little information is provided).

were systematically analysed, and how their results were derived. As a result, evaluators did not assess the quality and reliability of the sources which provided the main input in their evaluations. Here, it should be noted that the evaluators themselves were well aware of the non-significance (in statistical terms) of their results.

It is also worth noting that in some evaluations economic theory was used to predict the possible impact of operations, and to judge the relevance of their design. While this approach does not address the causality problem, it does improve the quality of the *a priori* evaluation of a project.

### ***Outcomes of evaluations: does aid for trade work?***

It has to be stressed that the reading of available evaluations does not offer a definitive answer to the question of whether – or why – aid for trade works for Ghana. This is because evaluators were unable to identify causal links between operations and performance. The results reported in evaluations regarding funds disbursed and projects implemented appear to be mixed. One recurring problem appears to be the operations' sustainability: several evaluations reported positive results in the short-term, but less positive results over the long-term because of financial uncertainty. Evaluators also mention the difficulty of tackling both growth and poverty reduction objectives simultaneously. For example, in one evaluation, support to the agricultural sector was said to have a pro-poor positive impact—by successfully targeting of poor recipients—but it was also said to have failed to promote growth or export diversification—raising serious questions about the sustainability of its pro-poor impact. Conversely, this same evaluation found that support for the transport sector had a positive impact on growth, but not on poverty reduction, as it was targeted mainly at the richest regions.

It is to be feared that reports of an operation's poor performance in terms of implementation also suggest that its impact on trade and poverty-reduction were limited at best. As a result, evaluations that lack instruments to identify causality probably also miss large positive results in terms of long-term impacts.

The nature of evaluations available for Ghana (mainly programmes or ‘aid policies’) probably helps explain the methodological limits outlined above. Indeed, such operations are difficult to assess given the complex factors impacting performance on a sectoral or macroeconomic scale, as opposed to the performance of a specific project in a specific context. The paucity of trade results might also reflect the absence of trade-related objectives in the operations' terms of reference.

### **Section 3. Aid for Trade in TranStor**

The *TranStor* sector is trade oriented by nature. All programmes or projects are devoted to rehabilitate, maintain or construct transport infrastructures (roads, rails or ports) and, at least indirectly, will normally lead to enhanced trade and economic growth, as well as reduced poverty. The World Bank (2007) provides a review of the literature on the importance of transport in the delivery chain, and the link between transport and poverty. Transport costs account for a large part of the value of imported and exported goods throughout the world. Especially in African or landlocked countries, transport costs seriously weaken the terms of trade of such countries<sup>7</sup>. The World Bank (2007) underscores the fact that transport effects on poverty reduction are not well understood but “growing

---

<sup>7</sup> According the World Bank (2007), 11.5 percent of the total value of imports to Africa is related to transport costs, and 20 percent of export costs and up to 55 percent for landlocked countries such as Malawi. There is an extensive literature on this matter, of which relevant examples are from Hummels (2007), Raballand, G., and P. Macchi (2008), Djankov *et al.* (2006).

*evidence links transport investment to the improved well-being of the poor*”.<sup>8</sup> Finally, although these effects are largely indirect, “most direct poverty-targeted interventions such as schools, clinics, or nutrition programmes depend on transport in one way or another”. In addition, in evaluations assessing the general background of the country in which a project operates, importance is given to the role of geography. For instance, landlocked or island countries are addressed differently given their unique geographical characteristics. These characteristics are carefully mentioned when designing the objective of the project.

### ***The set of selected evaluations***

Out of the 64 studies and evaluations covered by the DEReC database, twenty were selected because they had an international trade dimension (Table 3.2). The background of this narrow set of selected evaluations – as well as the 64 studies available in the DEReC database – is very similar. The proportions of donors in the overall and in the narrow set are quite comparable (with the AsDB counting for a majority of the evaluations). The same is observed for the projects’ implementation dates, which largely cover the period 1995-2000. Several evaluations refer to specific projects, while others refer to a set of projects (a programme).

**Table 3.2. Key data for the narrow set of evaluations: the TranStor sector**

ID	Implementation date	Period under scrutiny	Institution	Programme	Type	Product sector	or	Countries
1	1995-2005	1995-2007	WB	Programme	Review	?		
6	1996-2004	1996-2007	AfDB	Project	Evaluation	Road		Cameroon
7	1995-2001	2000-2005	ADB	Project	Evaluation	Road		Lao
9	1994-2001	1994-2003	WB	Project	Evaluation	road, bridge, rail		Croatia
12	1997-2005	1995-2007	ADB	Programme	Review	road, railway		China
13	1985-2005	1985-2005	ADB	Programme	Evaluation	road, bridge		Pakistan
17	1995-2001	1994-2005	ADB	Project	Evaluation	port, expressway		China
30	?	1997-2007	Denmark	Projects	Review	Road		
38	1991-1998	1989-2001	ADB	Project	Evaluation	road, bridge		Lao
42	1989-1998	1989-2001	ADB	Project	Evaluation	Port		Vanuatu
43	1992-1998	2000	ADB	Project	Evaluation	Port		Maldives
44	1988-1997	2001	ADB	Project	Evaluation	Port		India
49	1988-1998	2002	ADB	Project	Evaluation	Road		Indonesia
50	1993-2000	2002	ADB	Project	Evaluation	Airport		Mongolia
56	1992-2001	2004	WB	Project	Evaluation	Road		Sierra Leone
57	1981-2000	2000	AfDB	Programme	Review	Road		All five African regions
58	1986-1999	2004	AfDB	Project	Evaluation	Highway		Tanzania
61	1992-2003		Japan	Programme	Review	Road, bridge		Cambodia
62	1992-2008	2008	ADB	Programme	Review	Road, rail, airport		Mongolia
63	1987-2007	2007	ADB	Programme	Review	Road, railway		India

That said, most transport infrastructure projects pertain to the road sector. The diverse nature and quality of road sectors is well represented in the selection, as projects targeted both national and local

<sup>8</sup> The World Bank (2007) provides a series of examples through research projects and surveys. Other papers on this issue are provided by Stone, Susan & Strutt, Anna & Hertel, Thomas (2010).

roads, expressways, transit roads and even whole networks. The second main sector is the maritime sector, and projects focused on maintaining or constructing of ports (most of which, naturally, deal with international traffic). Projects focused on bridges, railways and airports were marginal in the narrow set of evaluations, as in the overall set. Finally, the narrow set covers a balanced range of recipient countries, even if Asian countries are over-represented in the sample.

#### ***Assessment of the Aid for Trade dimension: Methodologies and issues covered***

##### ***Methodologies***

An analysis of the selected evaluations leads to a number of general, as well as specific, observations. First, TranStor projects did not target specific populations. Although many projects impacted on rural populations, often in farming areas, this was not an objective of the projects *per se*. The same was true for the projects' impact on women.

Second, the project's success was always assessed at the project and/or overall level, especially in the case of a programme evaluation. As for country-related analysis, different dimensions were assessed, such as the projects' outcomes, relevance, effectiveness, efficiency, institutional impact, socio-economic impact, environmental impact, sustainability or cross-cutting impacts. Although the criteria used for each evaluation varied across donors, the more or less followed the Operating Evaluation Department's (OED) guidelines. Almost all AsDB's evaluations took the form of Project Performance Evaluation Reports (PPER), which followed the Project Completion Reports (PCR). Apart from citing the World Bank (2007) review, there were few references to other similar studies. Initial conditions were usually well identified in selected evaluations, with a clear description of both the background and the needs of the transport sector in question (best illustrated by the AsDB's evaluations). Despite describing initial conditions, evaluators often lamented the lack of useful baseline data to correctly validate the impact of projects.

Regarding methodologies, there were some disparities among evaluations. Delays and budgets were evaluated in all the selected studies. But while all evaluations attempted to provide some sort of rate of return (EIRR and ERR) assessment, only a few presented a cost/benefit analysis. Several evaluations tried to undertake a robust cost/benefit analysis by assessing alternative scenarios and providing sensitivity analysis.

Regarding outcomes, it should be emphasised that almost all the selected evaluations received a positive rating, in the sense that projects were usually found to be successful in meeting their initial objectives. Notwithstanding the potential selection bias involved in choosing evaluations from the DEReC database, the World Bank (2007) has pointed out that ratings in the transport sector have benefited from steady improvements since the early 1990s, and have scored better than other sectors. Yet despite satisfactory ratings, almost all projects suffered from delays (an average of 2 years) resulting from either internal project weaknesses or external factors. Surprisingly, most projects managed to stay within – or even come under – initial budgets. Where there were cost overruns, they seem to have been the result of exchange rate changes.

##### ***Covered issues: The impact on trade***

Only half of the selected evaluations explicitly focused on enhancing trade or facilitating the movement of goods as priority objectives. The other half focused on reducing transport costs, increasing economic growth, or improving living conditions. While international and domestic trade was not always explicitly included in the summary description of the project's objectives, trade was always mentioned by evaluators as one of the expected results of the project. However, when it came

to measuring trade impacts, figures were hard to come by and evaluators admitted their analyses' limitations.

As pointed out by the Nordic Consulting Group (2008), the universal indicators used in *TranStor* evaluations are Vehicle Operating Costs (VOC), Traffic Flows and Time Saving. For ports projects, the basic indicator is the port throughput, a series of efficiency indicators related to the productivity of the port. These indicators would have provided a useful basis for assessing the impact of transport improvements on trade<sup>9</sup>, but no extrapolation was done in the 20 evaluations covered. When a systematic measurement of transport costs proved impossible, some evaluations attempted to approximate the evolution of transport costs following the project's implementation.

Last but not least, it was generally the case that when trade was assessed; the assessment focused more on exports than imports. However, there were several notable exceptions, with some evaluators attempting to assess the project's likely impact on the cost of imports for the local population. This information was mostly gained through interviews.

#### *Covered issues: The impact on poverty*

It is important to stress that, even when poverty reduction was not cited in a project's summary objectives, many evaluations considered an operation's potential impact on this issue. Among the 20 selected evaluations, only two focused explicitly on the link between improved transportation and poverty reduction or improved living conditions, by examining the construction of schools, health centres, cultural centres, small agricultural projects, or water and electrification projects.

However, several other evaluations tried to assess this impact using a series of proxies or poverty indicators, such as school enrolment, access to health (average distance to a centre), the reduction in child illness, access to credit, women's participation in the economy, income evolution, and economic diversification. These evaluations also relied on socio-economic surveys and interviews, some of which were extremely detailed.

#### *Covered issues: Sustainability*

Most evaluations made reference to "sustainability", with the goal of assessing a project's impact on the environment and socio-economic issues. Under the umbrella of "sustainability" evaluators asked two questions: Were projects detrimental to the environment? And did projects provide new employments by strengthening the private sector?

Regarding environmental impacts, the evaluations were very general. The 20 selected evaluations generally concluded that the operations' environmental impact was neutral, but they offered little robust evidence to support this conclusion. One can assume that transport infrastructure improvements – particularly road improvements – resulted in increased traffic, yet the possibility that traffic growth would in turn lead to increased CO2 emissions (particularly in developing countries, where vehicle emission efficiency is generally lower) was never mentioned in the environmental assessments.

Regarding the private sector impact, evaluations largely acknowledged the key link between transport infrastructure improvements and private sector development. Moreover, private sector development was seen as central to poverty reduction in almost all evaluations. Several evaluations tried to measure projects' impact on business (especially small business) creation, industrial expansion

---

<sup>9</sup>As reminded previously transport cost is a substantial part of import and export value.

or even increased tourism. Increased employment – or, more interestingly, the net balance of jobs created and destroyed - was also measured in some evaluations.

### ***Outcomes of evaluations: does aid for trade work?***

The evaluations found that operations' trade impact was positive, based mostly on the growth of traffic volumes. However, these conclusions were reached without taking into account counterfactuals or the effects of other external factors. Attempts to measure operations' impact on poverty reduction suffered from the same weaknesses – although most evaluations saw poverty reduction as an indirect outcome and difficult to assess.

Beyond these common themes, some evaluations also recognised the synergies that can result from donors working together. Assessing these synergies - and other potential externalities - is difficult to quantify and, at a minimum, requires a certain level of agreement on methodologies among donors. This is especially true for a complex sector like transportation which requires coherence across modalities (road, railways, etc.) and territories to deliver the expected benefits. In particular, close attention needs to be paid to trans-modal coherence in any evaluation. For example, it makes little sense for a project to target just one aspect of transport infrastructure, say road maintenance, without reference to the broader transport network and regulatory environment of which it is a part. This concern is more adequately addressed in programmes evaluations that attempt to assess different transport-sector projects together and comprehensively.

### **Section 4. Final remarks**

DAC guidelines were followed relatively closely in most evaluations, in that projects were analysed in terms of their relevance, efficiency and sustainability, as well as in terms of their impact on a series of cross-cutting issues such as gender or the environment. However, the impact on trade was clearly not the focus of evaluators work. In a number of cases; it was not addressed at all. In other case, it was addressed only in very general terms, and the operations' impact on exporters and/or importers was hardly ever evaluated.

Some simple quantitative tools, such as rates of return, were sometimes used, especially when infrastructure projects were analyzed. However, except in few instances (such as in TranStor evaluations), more sophisticated tools, such as economic and econometric modelling, were rare, mainly because of a generalised lack of data collection, both *ex ante* (during operation implementation) and *ex post*.

Outcomes differed from one case to the next. While in Vietnam, all selected evaluations generally had positive ratings, in Ghana the results were mixed. However, this difference might stem from the very different nature of the operations under consideration. In both countries, donor harmonization and coherence with governments were usually mentioned and positively rated in the evaluations. In contrast, the operations' sustainability and impact were harder to document, and the outcomes seemed less satisfactory.

It appears that the quantitative (key word count) and the qualitative analyses point to similar conclusions - *e.g. the stock of Vietnam and TranStor provide a better selection of trade oriented projects and even evaluations*. Excluding Ghana, the word “export” was more often reported than “import”, as suggested above. Finally, the qualitative analysis confirms what the quantitative analysis showed – namely, that trade policy issues were rarely examined in the selected evaluations.

## CHAPTER 4. WHAT NEEDS TO BE DONE: LOOKING FOR A SIMPLE APPROACH

### Summary of the results

- There are huge variations among the evaluation procedures used. Moreover, economic or econometric techniques are uniformly ignored.

### Introduction

The previous chapters provided a rough picture of the selected evaluations, focusing on “what we want to know” and “what we effectively know”. This chapter focuses on “what we should do ideally” - and provides some recommendations. It prepares the ground by examining how current evaluations perceive the evaluation process. Then it suggests the kind of things that a model evaluation might look at.

### Section 1. Preparing the ground

It is interesting to look at the evaluation process as revealed in the evaluations themselves. This can be done in three steps: by looking at the evaluation procedures used; by looking at the economic instruments used for evaluating operations in other domains; and by looking at the econometric instruments used for evaluating in other domains.

#### *Evaluation procedures*

A frequent complaint of evaluators was that they lacked a robust *ex ante* assessment of the operation. Objectives tended to be defined without first undertaking the work required to make the operation “assessable”. In this respect, evaluations in the TranStor sector generally looked better than evaluations in Vietnam or Ghana because certain technical, engineering work had been done before launching an operation, and this preliminary work required a clear definition of objectives and means. It is revealing that the word “objective” appeared much more frequently than the word “outcome”.

Against this background, it is worth examining how the evaluators themselves perceive their own procedures. Table 4.1 presents the frequencies with which four “procedural” terms appeared in evaluations. The words “indicators” and “performances” relate to outcomes, while the words “monitoring” and “review” relate to procedures that are critical to ensuring the quality of aid for trade operations. Table 4.1 shows a notable variation in the frequencies of the four key words, with systematically lower frequencies for Vietnam and less systematically lower frequencies for TranStor.

**Table 4.1. Key evaluation procedures, by recipient**

	Overall set			Narrow set		
	Ghana	Vietnam	Transtor	Ghana	Vietnam	Transtor
Refer to indicator(s)?	20.9	13.9	61.2	15.3	17.0	16.3
Refer to performance(s)?	49.3	28.0	23.0	41.4	31.2	78.9
Refer to monitoring?	28.7	24.4	37.1	25.7	23.0	18.8
Refer to review?	55.0	44.4	18.7	47.2	41.6	38.1

Source: DEReC database. Authors' computations

By contrast, procedural words (except for indicators) appear more frequently in the narrow set of evaluations more closely related to the aid for trade agenda.

### **Key economic techniques**

Shifting from procedural to more substantive issues, Table 4.2 lists several terms that are part of routine evaluation work in other domains. The term “impact assessment” captures the idea of an integrated approach to public decision-making and is widely used in various national contexts—from the expected impact of future regulatory changes (Australia, the EC) to the *ex post* impact of regulatory changes already introduced (the U.S.). The terms “cost-efficiency” and “cost-benefit” analyses are well known techniques (with different strengths and weaknesses). Terms such as “short-term”, “long-term” and “discount rate” are also included in the list. Chapter 4 showed how taking into account adequate time horizons plays a key role in the accuracy and quality of evaluations, while the discount rate is crucial to cost-efficiency and cost-benefit analyses. Timescale issues are particularly important for developing countries which need to more carefully allocate limited resources over time.

Table 4.2 shows that all these terms were rarely mentioned by evaluators, with the exception of “time horizon”. In this respect, the adjective “long term” was mentioned more frequently than “short term” (suggesting that, on its own, “long-term” has little meaning without an awareness of the discount rate and its role in long-run cost benefit analysis) (Hahn 2000). Here the overall and narrow sets of evaluation do not differ greatly.

**Table 4.2. Key econometric techniques, by recipient**

	Overall set			Narrow set		
	Ghana	Vietnam	Transtor	Ghana	Vietnam	Transtor
Refer to impact assessment(s)?	1.6	0.7	3.0	2.5	3.8	2.0
Refer to cost-efficiency?	0.1	0.6	0.1	0.2	0.0	0.1
Refer to cost-benefit?	0.6	0.6	0.6	1.1	0.0	0.4
Refer to time horizon: short term?	5.6	4.0	2.2	7.2	2.5	2.0
Refer to time horizon: long term?	13.8	12.3	9.2	16.8	8.7	11.3
Refer to discount rate?	0.1	0.2	0.2	0.2	0.2	0.1

Source: DEReC database. Authors' computations.

### **Key economic techniques**

Other aid domains where public choice is involved (*e.g.*, education, health, etc.) increasingly use more sophisticated techniques (for an excellent survey, see World Bank 2009). In order to assess whether these techniques were employed by evaluations, Table 4.3 presents the frequencies with

which four key terms appeared in the selected studies. Almost all these terms were totally ignored in the evaluations in both the overall and narrow sets.

**Table 4.3. Key econometric techniques, by recipient**

	Overall set			Narrow set		
	Ghana	Vietnam	Transtor	Ghana	Vietnam	Transtor
Refer to counterfactual(s)?	1.1	1.2	0.3	0.8	0.0	0.3
Refer to control variable?	0.0	0.0	0.2	0.0	0.0	0.0
Refer to differences in differences?	0.0	0.0	0.0	0.0	0.0	0.0
Refer to randomization?	0.0	0.0	0.0	0.0	0.0	0.0

Source: DEREc database. Authors' computations.

## Section 2. Options for a better assessment of the Aid for Trade operations

As mentioned in the introduction, evaluations could be improved by using the economic methods captured in Table 4.2 or the more sophisticated methods captured in Table 4.3. A pragmatic approach would treat these two tracks as complementary. The advantage of economic methods is that they can be used *ex ante* as well as *ex post*—and the need of *ex ante* studies has been highlighted repeatedly by evaluators, not least because costs can be kept within reasonable limits (the cost of economic methods can also be significantly higher, depending the approach). In other words, economic methods can be routinely employed in any evaluation. Econometric methods are by definition more expensive (relying on micro-data collection) and require relatively higher skills to administer. In other words, they can be used in only a limited number of cases.

This paper focuses on the first track - which seems the most urgent – and highlights the importance of guidelines for evaluating aid-funded operations. However, a key lesson drawn from Chapters 2 and 3 is that there is no common pattern in the way donors evaluate their operations. This implies that current guidelines many not be precise enough, or may not be implemented strictly enough.

It is beyond the scope of this paper to review existing guidelines. It seems more useful to provide some information on economic techniques used relatively frequently in other domains – *e.g.*, in assessing the impact of future regulatory changes (in order to take into account the essential *ex ante* aspect).

In this context, Table 4.4 presents the simplified sequence of questions that the Australian Office of Best Practice Regulations asks evaluators to undertake when assessing the potential impact of regulatory changes.

**Table 4.4 Impact assessment: an ideal checklist**

<b>Step</b>	<b>1. Identifying the problem</b>
	1.1 Markets concerned ('relevant' markets)
	1.2 Market failure
	1.3 Regulatory failure
	1.4 Unacceptable risk
	1.4 Social goals
	1.5 Assessing the consequences of no action
<b>Step</b>	<b>2. Defining the objectives of government action</b>
	2.1 Defining the objective
	2.2 Defining the outcome
<b>Step</b>	<b>3. Examining the options that may achieve the objective / outcome</b>
	3.1 No action
	3.2 Market-based instruments
	3.3 Subsidies, taxes
	3.4 Is there a regulatory option?
	3.5 Are there other possible options
<b>Step</b>	<b>4. Impact analysis – costs, benefits and risks</b>
	4.1 Who is affected by the problem
	4.2 Who is likely to be affected by the proposed solutions?
	4.3 Costs <ul style="list-style-type: none"> <li>To producers</li> <li>To consumers</li> <li>To the community and / or environment</li> <li>To governments</li> </ul>
	4.4 Benefits <ul style="list-style-type: none"> <li>To producers</li> <li>To consumers</li> <li>To the community and / or environment</li> <li>To governments</li> </ul>
	4.5 Analysing risk
	4.5 A few key additional points <ul style="list-style-type: none"> <li>Competition assessment</li> <li>Effect on small businesses</li> <li>Effect on trade</li> <li>Ecologically sustainable development</li> <li>Deviation for international standards</li> </ul>
	4.6 Quantifying the impacts where significant <ul style="list-style-type: none"> <li>Valuing costs and benefits where there is no market</li> <li>Discounting</li> <li>Sensitive analysis</li> <li>Quantifying the compliance costs</li> </ul>
<b>Step</b>	<b>5. Consultation</b>
<b>Step</b>	<b>6. Conclusion and recommended option</b>
<b>Step</b>	<b>7. Implementation</b>
<b>Step</b>	<b>8. Review</b>

Source: Adapted from the Australian Office of Best Practice Regulation (2008)

Donors and recipients could adapt such a list to the aid for trade agenda. Such a modified list would not replace interviews or field information. In fact, ideally it would be combined with feedback from the people directly concerned with aid-funded operations. Nor would such a list eliminate the need for constant review, which serves to both assess the sustainability of operations over time, and discipline evaluators.

With these qualifications, several specific points need to be made about the “check list” outlined in Table 4.4. First, *ex ante* Steps 1 to 3 are essential and should be well developed – as underscored by the qualitative analysis above. Step 1 involving a substantial effort to determine whether the operation addresses a market or regulatory failure. Market failures cover a wide range of possibilities – from abuse of market power, to asymmetric information, to externalities, to public goods - which require different responses than regulatory failures - which in turn require a thorough analysis of the relevant regulations in place. Step 2 is about clarifying and defining the objectives—and ranking them according to primary and subsidiary outcomes. Too often an operation is treated as a single “stone” that can “kill two (or more) birds”. Step 3 is about examining all available options—from no action, to market-based instruments (tradable property rights, auctions, etc.), to subsidies, to taxes (with defined broadly so that they include international aid, of course), to regulatory reforms – and selecting the appropriate response

Step 4 focuses on impact analysis, and has two aspects. First, it should be done both *ex ante* – to provide evaluators with robust benchmarks - and *ex post*. Second, it should be interpreted broadly, in order to offer a comprehensive view of the operation. For instance, costs should include administrative costs, changes in sources of supply, changes in input prices, and so forth. Cost to consumers should not only include changes in the price of a good or a service, but also reductions in “utility” (e.g., quality, range of choice, etc.), undesirable income or wealth redistribution effects, unemployment, etc. Benefits to consumers should be looked at just as comprehensively. Sub-steps 4.4 to 4.6 are self-explanatory.

Finally, as noted above, Steps 5 to 8 deal with the inter-active process – *i.e.*, between the policy makers and the people directly affected by the operations. Here too the importance of the review process (Step 8) cannot be over-emphasised. If a review is done fairly, it can help to build trust between policy makers and people. And trust may well be the most crucial input for efficient aid for trade.

## ANNEX A. AN OVERVIEW OF THE EVALUATIONS, EVALUATORS AND THE TIME PATTERN

As underlined in the introduction, this meta-evaluation deals with only two countries - Ghana and Vietnam - and one sector - transport and storage, aggregated for the purposes of this study in the “TranStor” sector. As already mentioned, a detailed examination of the 162 evaluations strongly suggests that the “aid for trade” component is not equally strong for all evaluations. Thus, besides the “overall” set of 162 evaluations, a “narrow” set of 43 evaluations more clearly focused on aid for trade issues was selected. The criteria used for choosing these 43 evaluations is described in detail in Chapter 3. In short, within each country/sector-specific set of evaluations, the most “aid for trade-related” evaluations were selected. Hence, the narrow sets were defined in a manner which takes into account the specificities of the countries and the sector under scrutiny—they are thus not perfectly similar.

### Section 1. The stock of evaluations under scrutiny

Table A1 (bloc A) provides an overview of all the evaluations under scrutiny focusing on countries (no similar comparison could be made for TranStor since there is only one sector.) It makes a distinction between the overall and the narrow sets of evaluations. The overall set, which is the most important at this early stage of the meta-evaluation, suggests that operations in Vietnam were more intensely evaluated than those in Ghana.

**Table A1. An overview: the stock of evaluations (1999-2009)**

	Ghana			Vietnam			TranStor	
	overall	narrow	Period	overall	narrow	Period	overall	narrow
<b>A. Data on evaluations</b>								
Number of evaluations	34	13	1999-2009	64	9	1999-2009	64	20
Number of pages	3312	1858	1999-2009	5719	814	1999-2009	4227	1438
Average length of evaluations	97	143	1999-2009	89	90	1999-2009	66	72
<b>B. Trade data</b>								
Exports (current \$)	5.3		2008	62.7		2008	--	
Imports (current\$)	10.3		2008	80.7		2008	--	
Tariffs								
average applied tariff (%)	13.0		2008	16.8		2008	--	
average bound tariff (%)	92.5		2008	11.4		2008	--	
binding coverage (%)	14.3		2008	100.0		2008	--	
<b>C. Macroeconomic data</b>								
Population (million)	23.3		2008	86.2		2008	--	
GDP (current \$ bio.)	16.1		2008	90.7		2008	--	
GDP (at PPP \$ bio.)	33.9		2008	240.1		2008	--	
GDP/capita (PPP \$)	1454.9		2008	2785.3		2008	--	
Real GDP growth index	153.0	(2000=100)		179.0	(2000=100)		--	
Gross external debt (\$ bio.)	5.8	2008e		23.7	2008e		--	
Public debt (% GDP)	67.5	2008e		44.5	2008e		--	
Economic aid (\$ bio.)	1.65	2006-2009		2.95	2006-2009		--	
<b>D. Poverty data</b>								
Below poverty (% pop.)	28.5		2007	7.8		2009e	--	
Gini index	39.4		2005	37.0		2004	--	

Source: DEREc, WTO. Authors' computations. For Vietnam, the evaluation done in 2010 included in the qualitative approach (Chapter 3) is not included in this table.

These observations need to be put in perspective by taking into account the basic differences between these two economies. Such differences can be defined in three different ways. First, they could be defined in terms of the relative size of trade flows, a definition that echoes the pure “trade” aspect of aid for trade. As Ghana’s exports and imports are much smaller (one fourth to one seventh) than those of Vietnam, Ghana emerges as much more monitored than Vietnam (Table A1, bloc B). Second, the differences between the two economies could be captured by the relative size of populations, GDP or other key macroeconomic variables (debt and economic aid). All these dimensions shift the focus from the pure “trade” aspect of aid for trade to indicators more directly related to development and aid. As Ghana’s macroeconomic indicators are still smaller than Vietnam’s (roughly one third to one fourth), Ghana continues to be more intensively monitored than Vietnam in this context—on average two to three times more intensely (Table A1, bloc C).

The third possible definition of the differences between the two countries provides a different perspective. This definition focuses on income distribution, defined as the share of the population below poverty level or the Gini coefficient. As Ghana exhibits higher income distribution than Vietnam, Vietnam appears now as more frequently subjected to evaluations than Ghana (Table A1, bloc D). This result raises interesting questions—once again assuming that the sample of evaluations is representative. Does it reflect the different nature of the operations evaluated, namely projects, programmes and general aid policy (see section 2 below)? Or does it mirror countries specificities (it may be easier to undertake operations in a country like Vietnam than in a country like Ghana)? Or does it reflect the fact that evaluators are more sensitive to key development concerns, such as income distribution, if a country has already reached a certain level of development, and seems on a successful growth path?

## Section 2. Different types of operations

Before going further, it is important to realise that the nature of the operations evaluated varies considerably, and that this feature makes the operations more or less easily subjected to evaluations. Table A2 presents a breakdown of the evaluations under scrutiny into three main categories, depending on the type of operations evaluated. Following the definitions provided by the OECD [1992], projects involve “*well-defined development activities that are limited in time and space*” while programmes “*consist of all contributions made available to a recipient country for general development purposes, i.e. balance-of-payments support, general budget support and commodity assistance, not linked to specific project activities*”. We further define “general aid policy” as the complete body of operations by a specific donor in a specific recipient country in a particular time frame.

Intuitively, the different scope of these three types of operations has an impact on the characteristics of the evaluations. Evaluating projects seems less easy to relate to broad trade and development issues than programmes and general policy operations for a host of reasons. Projects are also likely to be more limited in terms of time horizon (i.e., to be shorter term), and hence may not allow sufficient time for assessing consequences. Conversely, evaluating the achievements of projects might be methodologically easier than those of a general aid policy, with much larger and long-term goals.

**Table A2. A breakdown by project, programme and general aid policy**

	Ghana		Vietnam		Transtor	
	overall	narrow	overall	narrow	overall	narrow
<b>A. All evaluations</b>						
Number of evaluations	34	13	64	9	64	20
Number of pages	3312	1858	5719	814	4239	1438
Average length of evaluations	97	143	89	90	66	72
<b>B. Evaluations of projects</b>						
Number of evaluations	11.0	2.0	27	5	38	12
Number of pages	670	114	2143	483	1949	566
Average length of evaluations	61	57	79	97	51	47
<b>C. Evaluations of programs</b>						
Number of evaluations	12	4	18	4	18	8
Number of pages	1153	485	1675	370	1845	872
Average length of evaluations	96	121	93	93	103	109
<b>D. Evaluations of general aid policy operations</b>						
Number of evaluations	11	7	19	--	8	0
Number of pages	1489	1259	1901	--	504	--
Average length of evaluations	135	180	100	--	63	--

Source: DEReC. Authors' computations. For Vietnam, the evaluation done in 2010 included in the qualitative approach (chapter 3) is not included in this table.

### Section 3. The time pattern of the evaluations by recipient

This section focuses on the time pattern of the evaluations under scrutiny according to recipient, by looking first at the overall set, and then at the narrow set, of evaluations.

#### *The overall set of evaluations*

Table A3 covers the whole set of evaluations. It shows a very different time pattern across countries and sectors. The evaluations of operations undertaken in Ghana—whether measured by the number of evaluations, the number of pages, or the average length of evaluations—are much more volatile than those made in Vietnam, even if, for both countries, the peak of evaluations occurred in the years 2004-2006. This volatility can be measured by an indicator of variance. The chosen indicator is the weighted standard deviation which captures how far each observation lies from the mean (the lower such an indicator is, the lower the volatility is). This indicator of variance is two to four times higher for Ghana than for Vietnam. Evaluations on the TranStor sector suggest a time pattern closer from the one in Ghana than the one in Vietnam.

That said, it is essential to go beyond the mere numbers of evaluations or pages, and to provide a measure reflecting the potential informational content of the evaluations. As explained in the methodology section of chapter 1 (and explained in more details in Annex A), the frequency of the 48 key words or expressions (the occurrence of such words per one hundred pages) provides such a—admittedly crude—measure of this potential informational content of the evaluations.

Table A3 gives the frequency by recipient for the overall set of the 48 key words or expressions. The average frequency of the key words and expressions is notably lower for Ghana than for Vietnam and TranStor. Interestingly, the variance is higher for Ghana than for Vietnam and TranStor.

**Table A3. A overview of the overall set of evaluations: the time pattern by recipient**

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	All years	Weight st.dev
<b>Ghana</b>													
Number of evaluations	1	3	0	1	0	7	8	8	2	2	2	34	1.00
Number of pages	528	402	--	109	--	493	432	851	205	142	150	3312	0.66
Avg length of evaluations	528	134	--	109	--	70	54	106	103	71	75	97	1.07
Frequency [a]	546	801	--	693	--	413	432	737	865	1403	616	730	0.41
Annual GNI (current \$)						460	530	590	630	n.a.	--		
<b>Vietnam</b>													
Number of evaluations	2	4	7	3	5	8	8	9	6	6	6	64	0.38
Number of pages	506	336	506	226	247	564	744	1139	435	679	337	5719	0.51
Avg length of evaluations	253	84	72	75	49	71	93	127	73	113	56	89	0.58
Frequency [a]	353	574	460	506	706	706	824	844	643	938	772	688	0.27
Annual GNI (current \$)						620	690	770	890				
<b>Transtor [b]</b>													
Number of evaluations	2	11	3	11	1	12	6	7	7	4	0	64	0.72
Number of pages	100	1044	112	549	100	498	302	466	812	244	--	4227	0.75
Avg length of evaluations	50	95	37	50	100	42	50	67	116	61	--	66	0.41
Frequency [a]	611	587	549	515	690	644	634	642	878	930	--	671	0.20

**Note:** [a] See Annex B for the list of words and expressions

Source: DEReC. Authors' computations. For Vietnam, the evaluation done in 2010 included in the qualitative approach (Chapter 3) is not included in this table.

These observations raise the following question: Are such differences explained by reasons related to the recipient countries or by reasons related to the donors? Part of the answer may be related to the recipient.

In the case of Ghana, “at-the-border” policies were subjected to two forces during this decade [Trade Policy Review, 2001 and 2008]. On the one hand, the simple average applied tariff fell from 17 percent in 1992 to 14.7 percent in 2000 and 12.7 percent in 2007. On the other hand, Ghana’s tariffs are still largely unbound in the WTO, with a very low binding coverage of 14.5 per cent (see Table A1), covering comprehensively, but practically only, agricultural products. This means that, at any time, tariffs could be significantly with little risk of retaliation by trade partners. For instance, in February 2000 Ghana introduced a “special import tax” of 40 per cent on a significant share of consumer goods [Trade Policy Review, 2001].

What is more, since 2009, the world consensus on open trade policies has come under stress because of the global financial crisis and recession. As just noted, because of its low bindings coverage, Ghana retains the option of increasing its applied tariffs at any time without risk breaking its WTO commitments or being forced to provide compensations to its trading partners. The potential reversal of its relatively open trade policy would be easy and quick. This situation is reinforced by Ghana’s macroeconomic imbalances which leave open the possibility of threatened or real trade policy reversals [Trade policy Review 2001].

In the case of Vietnam, the trajectory of its “at-the-border” and “behind-the-border” policies is much more ambitiously liberal (this is partly due to the fact that Vietnam was largely a closed economy in the 1980s) - and appears firmer - than Ghana’s. In trade matters, the decade under consideration was dominated Vietnam’s WTO accession negotiations. Vietnam’s Protocol of

Accession was ratified in 2007, after a 12 year negotiations (a relatively short period of time). The Protocol includes commitments on a wide range of both “at-the-border” and “behind-the-border” policies. Vietnam has agreed to bind all its tariffs at an average rate of 11.4 per cent (see Table A1), and to undertake ambitious disciplines on non-tariff barriers (from quotas to standards), foreign direct investment, services, and intellectual property. This is in sharp contrast to Ghana which has no commitment in many policy areas and sectors. The extremely wide ranging commitments adopted under the Protocol have served to reduce fears among donors of policy reversals, even though some of these commitments are being implemented at a relatively slow pace.

### ***The narrow set of evaluations***

As noted above, the overall set of evaluations is heterogeneous, with many evaluations being only loosely related to aid for trade. As a result, Table A4 is limited to the narrow set of 43 evaluations which can be considered most robustly linked to aid for trade.

**Table A4. An overview of the narrow set of evaluations: the time pattern by recipient**

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	All years	Weight st.dev
<b>Ghana</b>													
Number of evaluations	1	3	0	0	0	0	3	4	0	1	1	13	1.24
Number of pages	528	402	--	--	--	--	205	561	--	104	58	1858	0.70
Avg length of evaluations	528	134	--	--	--	--	68	140	--	104	58	143	1.03
Frequency [a]	546	801	--	--	--	--	1061	736	--	1513	336	763	0.50
<b>Vietnam</b>													
Number of evaluations	0	0	0	0	1	0	0	2	1	2	3	9	1.32
Number of pages	--	--	--	--	46	--	--	220	51	303	194	814	0.69
Avg length of evaluations	--	--	--	--	46	--	--	110	51	152	65	90	0.53
Frequency [a]	--	--	--	--	561	--	--	1274	629	1369	828	1134	0.40
<b>Transtor</b>													
Number of evaluations	0	2	2	4	0	3	1	2	3	3	0	20	0.77
Number of pages	--	110	92	162	--	97	76	160	551	190	--	1438	0.86
Avg length of evaluations	--	55	46	41	--	32	76	80	184	63	--	72	0.67
Frequency [a]	--	634	501	607	--	787	445	564	1000	1019	--	819	0.32

**Note:** [a] See Annex B for the list of words and expressions

Source: DEReC. Authors' computations.

### **Section 4. The evaluations by donor**

This section switches the focus from recipients to donors. It excludes four evaluations from the sets of evaluations because the corresponding donors (IDO in the case of Ghana, EBRD, NEI and NORA in the case of TranStor) have only one evaluation for the countries and sectors under scrutiny. Once again, a distinction is made between the overall set of evaluations and the narrow, more aid for trade-focused set of evaluations.

### ***The overall set of evaluations***

Table A5 suggests several conclusions. First, each donor's share in terms of the number of evaluation pages gives a rough sense of the structure of the “supply” side of the “evaluation market”.<sup>10</sup> From this perspective, there is a marked difference between Vietnam and TranStor, on the one hand,

<sup>10</sup> The number of pages seems a more accurate indicator than the number of evaluations.

and Ghana, on the other. Evaluations of Vietnam and TranStor are largely dominated by one or two donors: SIDA and, to a smaller extent, AsDB for Vietnam (totally 75.9 per cent of all the pages); AsDB and (to a notably smaller extent) AfDB and the World Bank (altogether 76.5 percent of all the pages) for TranStor. By contrast, the Ghana evaluations reveal no dominant donor, with four donors (AfDB, DANIDA, DFID and the World Bank) having almost the same share of pages (15 to 17 per cent), although AfDB remains an important provider in terms of the number of evaluations.

This difference between a supply structure dominated by a couple of donors and a supply structure fragmented among several donors is an interesting feature – and might warrant further investigation as to whether these differing supply structures contributed to the big dispersion in evaluations for Ghana compared to those for Vietnam. That said, it should be stressed that this observation is based on evaluations notified to DEReC, not on all the evaluations undertaken by donors in these two countries. Thus, it may give a distorted view of the situation.

**Table A5. An overview of the overall set of evaluations: the donor pattern**

	AfDB	AsDB	AFD	CIDA	DANIDA	DFID	EC	GTZ	Japan	SECO	SIDA	USAID	WB	J-Eval	ALL	Weighted std.dev
<b>Ghana</b>																
Number of evaluations	9	--	1	1	4	5	1	--	1	--	1	4	6	--	33	0.85
Number of pages																
total	322	--	58	38	600	690	124	--	26	--	104	201	621	--	2784	0.94
share (%)	11.6	--	2.1	1.4	21.6	24.8	4.5	--	0.9	--	3.7	7.2	22.3	--	100.0	0.94
Avg length of evaluations	36	--	58	38	150	138	124	--	26	--	104	50	104	--	84	0.56
Frequency of key words and expressions [a]																
complete list	934	--	352	1132	900	779	1167	--	812	--	513	487	715	--	868	0.35
coreP list	397	--	72	587	362	382	373	--	419	--	191	149	384	--	336	0.46
coreP loss rate	42.5	--	20.5	51.9	40.2	49.1	32.0	--	51.6	--	37.2	30.6	53.7	--	38.7	0.27
<b>Vietnam</b>																
Number of evaluations	--	19	2	1	--	1	2	2	1	3	28	--	2	3	64	1.55
Number of pages																
total	--	1790	80	34	--	83	162	120	13	130	2721	--	153	433	5719	1.71
share (%)	--	31.3	1.4	0.6	--	1.5	2.8	2.1	0.2	2.3	47.6	--	2.7	7.6	100.0	1.71
Avg length of evaluations	--	94	40	34	--	83	81	60	13	43	97	--	77	144	89	0.53
Frequency of key words and expressions [a]																
complete list	--	826	358	1035	--	906	1515	600	708	987	484	--	827	961	687	0.37
coreP list	--	340	44	538	--	439	386	273	277	268	161	--	348	425	257	0.43
coreP loss rate	--	41.2	12.3	52.0	--	48.5	25.5	45.5	39.1	27.2	33.3	--	42.1	44.2	37.4	0.31
<b>Transtor</b>																
Number of evaluations	9	35	--	--	3	--	--	1	1	--	3	--	9	--	61	1.39
Number of pages																
total	485	2179	--	--	203	--	--	6	19	--	290	--	555	--	3737	1.41
share (%)	13.0	58.3	--	--	5.4	--	--	0.2	0.5	--	7.8	--	14.9	--	100.0	1.41
Avg length of evaluations	54	62	--	--	68	--	--	6	19	--	97	--	62	--	61	0.58
Frequency of key words and expressions [a]																
complete list	782	642	--	--	643	--	--	850	716	--	505	--	801	--	674	0.17
coreP list	416	293	--	--	293	--	--	283	126	--	169	--	367	--	295	0.37
coreP loss rate	53.2	45.6	--	--	45.6	--	--	33.3	17.6	--	33.5	--	45.8	--	43.8	0.30

**Note:** [a] See Annex B for the list of words and expressions

Source: DEReC. Authors' computations.

Second, Table A5 tries to give a sense of the potential informational content of the evaluations by donor, by showing the frequency with which all the words and expressions defined as key appeared. It reveals a wide gap—by a factor of four for Ghana, or six for Vietnam—among the donors exhibiting the highest and lowest frequencies. The relative ranking of all the donors (for instance, those with frequencies higher than the average frequency for the overall stock of evaluations) is relatively stable. The same observation is true if one limits the donors presented to those with the highest number of evaluations. This stability of results is interesting, since it suggests some systemic factors—but, it would need to be tested on a higher number of evaluations for the donors with only a few evaluations

in order to be confirmed. In sharp contrast to the countries cases, the frequency gap between donors is much smaller in the TranStor case, with a factor of only 1.5.

These results rely on the overall set of the 48 key words or expressions. It seems useful to refine this list by focusing on subsets of words and expressions considered particularly important (“core”). Such a core list is defined by the 30 words and expressions which are considered by the authors as truly essential from a purely economic perspective. In other words, it gives a sense of the evaluations which are the most clearly driven by economic analysis. The core list of words and expressions is provided in Annex A. It consists in three groups (of roughly similar size) of words and expression which are related to trade, development and options for future guidance.

The core list shows much smaller frequencies—revealing a much smaller potential informational content on pure economic issues. The “loss rate” (the number of economic-oriented words as a ratio of the words included in the overall list) is substantial—on average 38 to 44 per cent. Moreover, it varies considerably among donors. It can reach up to 80 per cent, and it is systematically higher than the average for a few donors (AFD, EC, SIDA and USAID). Finally, as a result of these different loss rates, a new stable ranking emerges among all the donors, with CIDA exhibiting the highest frequencies for Ghana and Vietnam, and AfDB for the TranStor sector.

### ***The narrow set of evaluations***

The narrow set relies on a substantially smaller number of observations (including in terms of donors). This feature means that the observations need to be interpreted with caution. However, it seems clear that there are one or two dominant donors in this set of evaluations, and that the wide gap in terms of frequencies among donors still prevails.

**Table A6. An overview of the narrow set of evaluations: the donor pattern**

	AfDB	AsDB	AFD	CIDA	DANIDA	DFID	EC	GTZ	Japan	SECO	SIDA	USAID	WB	J-Eval	All donors	Weighted st.dev	
<b>Ghana</b>																	
Number of evaluations	4	--	1	--	3	2	1	--	--	--	--	--	--	1	--	12	0.63
Number of pages																	
total	189	--	58	--	508	407	124	--	--	--	--	--	--	44	--	1330	0.87
share (%)	14.2	--	4.4	--	38.2	30.6	9.3	--	--	--	--	--	--	3.3	--	100.0	0.87
Avg length of evaluations	47	--	58	--	169	204	124	--	--	--	--	--	--	44	--	111	0.64
Frequency [a]	886	--	336	--	920	685	1167	--	--	--	--	--	--	1184	--	849	0.37
<b>Vietnam</b>																	
Number of evaluations	--	3	--	--	--	--	1	1	--	3	2	--	--	--	--	10	0.50
Number of pages																	
total	--	392	--	--	--	--	99	115	--	130	78	--	--	--	--	814	0.80
share (%)	--	48.2	--	--	--	--	12.2	14.1	--	16.0	9.6	--	--	--	--	100.0	0.80
Avg length of evaluations	--	131	--	--	--	--	99	115	--	43	39	--	--	--	--	81	0.49
Frequency [a]	--	1260	--	--	--	--	1783	580	--	987	736	--	--	--	--	1134	0.44
<b>Transtor</b>																	
Number of evaluations	12	3	--	--	1	--	--	--	1	--	--	--	3	--	20	1.15	
Number of pages																	
total	942	160	--	--	47	--	--	--	19	--	--	--	270	--	1438	1.32	
share (%)	65.5	11.1	--	--	3.3	--	--	--	1.3	--	--	--	18.8	--	100.0	1.32	
Avg length of evaluations	79	53	--	--	47	--	--	--	19	--	--	--	90	--	72	0.48	
Frequency [a]	791	766	--	--	847	--	--	--	716	--	--	--	953	--	819	0.11	

**Note:** [a] See Annex B for the list of words and expressions.

Source: DEReC. Authors' computations.

## ANNEX B. A BRIEF PRESENTATION OF THE LIST OF KEY WORDS

Annex B provides more information on the methodology used in the paper. Table B1 presents the complete list of words and expressions that correspond to issues on which, ideally, policy makers would like to get more information from field evaluators in order to improve the gains from development aid. For instance, when the question “is trade taken into consideration?” is asked, the answer is provided by the number of occurrences of the word “trade” in the evaluations. These words are divided in three columns: column 1 lists words related to trade aspects, column 2 those related to the development dimensions, and column 3 those related to options for future guidance. There is a total of 48words and expressions, almost equally split among these three aspects.

**Table B1. The complete list of words and expressions**

Trade component	Table	Development component	Table	Procedures and techniques	Table
trade	2.1	infrastructure	2.5	indicator	4.1
export	2.1	specific regulation	2.5	performance	4.1
import	2.1	regulatory framework	2.5	monitoring	4.1
trade balance	2.1	governance	2.5	review	4.1
comparative advantage	2.1	supply-side constraint	2.6	impact assessment	4.2
gains from trade	2.1	expenditure	2.6	cost efficiency	4.2
trade restriction	2.2	private sector growth	2.6	cost-benefit	4.2
tariff	2.2	technical assistance	2.6	short term	4.2
quota	2.2	economic growth	2.7	long term	4.2
subsidy	2.2	competitiveness	2.7	discount rate	4.2
technical barriers to trade	2.3	efficiency	2.7	counterfactual	4.3
sanitary/phytosanitary standards	2.3	effectiveness	2.7	control variables	4.3
trade facilitation	2.3	sustainability	2.7	difference in differences	4.3
adjustment policies	2.3	poverty reduction	2.7	randomization	4.3
trade assistance	2.3	gender	2.7		
trade-related technical assistance	2.3				
WTO	2.4				
regional trade agreements	2.4				
preferential trade agreements	2.4				

Section 1 shows that it is very useful to defined “core” words. The core list focuses on words and expressions key in the context of economic analysis, as defined by the authors. It is beyond the scope of this paper to explain the choice of the core list in detail. And, there are several reasons for not including a word in the core list. For instance, the word “trade” is too wide either because it refers to realities indirectly important from the international trade perspective on which this paper focuses (for instance, it can refer to retail trade) or because it is not precise enough to capture the essence of the economic gains from trade (too much trade can be as harmful as too little trade). Table B2 provides the list of these words.

**Table B2. The core list of words and expressions key from an economic perspective**

Trade component	Table	Development component	Table	Procedures and techniques	Table
import	2.1	specific regulation	2.5	indicator	4.1
comparative advantage	2.1	regulatory framework	2.5	performance	4.1
gains from trade	2.1	supply-side constraint	2.6	monitoring	4.1
tariff	2.2	economic growth	2.7	review	4.1
quota	2.2	competitiveness	2.7	impact assessment	4.2
subsidy	2.2	efficiency	2.7	cost efficiency	4.2
technical barriers to trade	2.3	effectiveness	2.7	cost-benefit	4.2
sanitary/phytosanitary standards	2.3	sustainability	2.7	discount rate	4.2
trade facilitation	2.3	poverty reduction	2.7		
adjustment policies	2.3	gender	2.7		

## BIBLIOGRAPHY

Djankov, S. & Freund, C. & Pham, C. (2006), “Trading on time”, Policy Research Working Paper Series 3909, The World Bank.

Hummels, D. (2007), “Transportation costs and international trade in the second Era of globalization”, *Journal of Economic Perspectives*, Volume 21, Number 3, Summer 2007, Pages 131–154.

Nordic Consulting Group (2008), “Synthesis Report, Overview and Synthesis of Road Sector Evaluations”, February 2008.

OECD (2006), *Aid for Trade – Making it Effective*, The Development Dimension. ,OECD, Paris.

Raballand, G., and P. Macchi (2008), *Transport Prices and Costs: The Need to Revisit Donors' Policies in Transport in Africa*, The World Bank, Washington DC.

Stiglitz, J. and A. Charlton, (2005), “Aid for Trade: A report for the Commonwealth Secretariat”, March, Commonwealth Secretariat.

Stone, S., A. Strutt, and T. Hertel (2010), “Assessing Socioeconomic Impacts of Transport Infrastructure Projects in the Greater Mekong Subregion”, Asian Development Bank Institute, ADBI Working Papers 234.

World Bank (2007), *A Decade of Action in Transport - An Evaluation of World Bank Assistance to the Transport Sector, 1995–2005*, The World Bank, Washington DC

Zedillo, E. and P. Messerlin (2005), *Trade for development – Achieving the Millennium Development Goals*, United Nations Development Program Publication, Earthscan Publishing.