STRENGTHENING ACCOUNTABILITY IN AID FOR TRADE

ADVANCE COPY

Corrigenda to OECD publications may be found on line at: *www.oecd.org/publishing/corrigenda*

Foreword

The Paris Declaration on Aid Effectiveness (2005) orients the aid relationship towards genuine partnerships. Mutual accountability is a concept designed to focus these partnerships on delivering results. Three elements are central in establishing mutual accountability. The first is a shared agenda, with clear objectives and reciprocal commitments. The second involves monitoring and evaluating these commitments and actions. Both of these elements inter-relate with the third: dialogue and review.

The Aid for Trade Initiative is one of the clearest international examples of how these three elements create powerful incentives to carry out commitments and, ultimately, to change behaviour. Accountability between partners and donors is enhanced by transparency about the effectiveness of aid programmes and projects and about learning what works and what does not work.

In that context, evaluations should look beyond whether intended outcomes have been achieved. They should also provide forward looking recommendations on how to make aid programmes more results-oriented. Despite the momentum towards introducing a true performance culture in development assistance, measuring results remains intrinsically difficult. For instance, the 2011 Survey on Monitoring the Paris Declaration shows that while some progress has been achieved in a number of countries where sound, results-based monitoring frameworks have been established, the pace is still too slow.

Strengthening Accountability in Aid for Trade shows that assessments of the effectiveness of programmes and projects are constrained by incomplete reporting on aid-for-trade outcomes. Consequently, institutional learning about what works best in building trade capacities is inhibited. The poor state of evaluation in aid for trade is not just a case of poor data entry, missing files, and still confusing cross-cutting aid categories. The difficulty in measuring outcomes points to systematic problems at the design and implementation stage, and in assigning macroeconomic outcomes and impacts to individual aid-for-trade programmes and projects.

This report suggests that improvements can be achieved through introducing a focus on results throughout the programme cycle, by means of independent joint evaluations and a greater focus on systematic learning. In addition, it acknowledges the problems and inconsistency that would occur if individual donors and development institutions were to pursue the results agenda on their own. Some suggest putting the emphasis on agreeing on a limited, simplified and unified set of indicators. Others caution that context specificity should be retained, and that care should be taken not to re-impose implicit conditions through a rigid, centralised set of results indicators. *Strengthening Accountability in Aid for Trade* argues that a joint but differentiated approach to working towards a menu of indicators for measuring results would be useful. Obviously, the active engagement of developing countries is a prerequisite in developing such a menu of indicators.

Brian Atwood
Chair of the Development
Assistance Committee

Acknowledgments

Strengthening Accountability in Aid for Trade was prepared under the auspices of the OECD Development Assistance Committee (DAC) and the Trade Committee. The report was written by William Hynes and Masato Hayashikawa of the OECD Development Co-operation Directorate, under the guidance of Frans Lammersen. The authors acknowledge the valuable contributions from a team of consultants, comprised of Claire Delpeuch, Marie-Agnès Jouanjean, Alexandre Le Vernoy and Thomas Orliac, and led by Professor Patrick Messerlin of SciencesPo Paris, which form the basis of Chapters 2 and 3. The authors are grateful to John Smith for editing the report, to Jean-Jacques Hallaert, Anna Hellstrom, Halla Hrund Logadóttir, Megan Kennedy-Chouane, Hans Lundgren, Bill Nicol, Michael Plummer, and the participants of the OECD *Experts Meeting on Indicators: Measuring Aid-for-Trade Results at the Country Level*, for their valuable inputs and comments on earlier drafts. Valuable assistance was provided by Susan Hodgson.

TABLE OF CONTENTS

Executive Summary

At a time when aid budgets are under pressure and scrutiny, there is a need to improve accountability. This is especially true in the case of aid for trade, which has become an increasingly important priority in development co-operation. However, the promise and possibilities of improved accountability should be kept in perspective. For instance, evaluation concepts and tools have proliferated in recent years, but in the opinion of some critics this has created a "fixation with measurement" or an "audit culture", accompanied by "bulging toolboxes loaded with frameworks and concepts".[1]

This report looks at what the trade and development community needs to know about aid for trade results, what past evaluations of programmes and projects reveal about trade outcomes and impacts, and how the trade and development community could improve the performance of aid for trade interventions.

Although donors have been evaluating trade-related assistance for many years, it is not easy to measure the outcomes and impacts of aid for trade. A number of conceptual difficulties, described in Chapter 1, need to be acknowledged and (where possible) overcome. A range of existing techniques and guidelines can be used to do this, and many of the building blocks are already in place. Careful planning of evaluations is crucial and should be undertaken from the beginning of a programme or project. A combination of *exante* reviews to determine what a programme or project is intended to achieve, and *ex post* assessments to determine what works and what does not, would be useful in advancing work on evaluation and would help develop synergies with work on managing for development results (MfDR).

A recent meta-evaluation by the OECD of 162 trade-related evaluations, analysed in Chapters 2 and 3 and Annex A, provides an overview of (and perspective on) the way OECD Development Assistance Committee (DAC) donors and international agencies have implemented programmes and projects and conducted evaluations in terms of both the methods used and topics covered. The OECD meta-evaluation demonstrates that evaluations of aid-for-trade programmes and projects have not usually had much to say about trade, and have had even less to say about the policy linkages that matter most to policy makers. They have referred extensively to broad, development-related concepts such as gender or poverty reduction, but without clearly defining these concepts. Moreover, evaluations have often lacked an adequate or realistic timeframe for measuring results, rarely distinguishing between what is achievable in the short and longer terms. Consequently, their conclusions give little insight into whether aid for trade works or why. The findings of the OECD meta-evaluation also need to be put into perspective. Evaluations' failure to refer to specific trade results can be explained, at least in part, by the absence of trade-related objectives in the initial mandate of programmes and projects.

The 2005 Paris Declaration on Aid Effectiveness put the results agenda firmly at the centre of global efforts to improve aid effectiveness. Against this backdrop, donors are increasingly using a results-based management framework to ensure that their activities

achieve the desired objectives and targets. The challenge of designing effective results chains that connect individual project objectives with more strategic, long-term development outcomes is addressed in Chapter 4. In measuring results at the outcome level, donors and partners alike are often confronted by the problem of attribution, *i.e.* what part of the observed changes resulted from aid-for-trade activities at the programme or project level?

Targets and results are specific to individual programmes and projects. The associated indicators also vary. Still, there are (or should be) commonalities between sector or macro level outcomes that can be quantified and aggregated into summary indices for benchmarking and cross-country comparisons. Although many indicators related to aid for trade have been developed in recent years, a reduction in the number of such indicators would help practitioners aggregate results data systematically across programmes and projects at the country, regional and global levels. Enhanced transparency would, in turn, contribute to a broader effort to make aid for trade more effective.

Notes

[1] Quoted from a report on a "reimagining development" seminar in 2010 organised by the Institute for Development Studies at the University of Sussex, United Kingdom. See the blog "From poverty to power: how effective citizens and effective states can change the world" by Duncan Green, Oxfam International (*www.oxfamblogs.org/fp2p/?s=Sussex*).

Chapter 1
How to Evaluate Aid for Trade: Approaches, Methods and Processes

"Not everything that can be counted counts, and not everything that counts can be counted."
– Albert Einstein

1.1 Introduction

This chapter looks at what the aid-for-trade community needs to know in order to evaluate aid for trade. No universal approaches exist for use in evaluating the broad and diverse range of activities that are part of the Aid for Trade Initiative. Donors use different definitions and approaches. They also have different objectives and resources. Thus, common evaluation frameworks are not required. In addition, it is clear that the evaluation community does not feel that new guidelines are needed for aid for trade: generic approaches are considered sufficient. Guidelines specifically tailored to different areas have proliferated in recent years, but they provide little added value. On the other hand, a discussion of conceptual challenges – particularly linkages in results chains, the problem of attribution, and the sequencing of programmes and projects – could result in greater operational clarity.

Aid for trade is not new. Evaluations have been undertaken regularly as part of donor activities, but in areas with a potentially important trade dimension that dimension has not always been considered. This is partly because of trade's low profile as an objective of development co-operation, but also because of methodological challenges. Yet there is an urgent need for new knowledge about aid-for-trade interventions. Evaluations (*ex post* assessments of effectiveness and impact) that assess the extent to which programmes and projects have met expectations, and that measure impacts, are essential to obtain evidence and insights.

Against this background, it is recommended that better evaluative guidance be provided to policy and management staff in both the *ex ante* formulation of aid-for-trade programmes and projects and their implementation. This chapter provides an overview of OECD work on approaches, methods and processes for the evaluation of aid for trade. It also explores the results chains some donors have developed for transport infrastructure and whether impact assessments are applicable in this area. A number of processes which would enable better evaluation are noted, such as independent, joint evaluations conducted with the participation of partner country representatives.

Since the World Trade Organization (WTO) Task Force on Aid for Trade met in 2006, there have been repeated calls for rigorous aid-for-trade evaluations. The Task Force outlined the potential objectives of evaluation, which included building knowledge and facilitating a results-based approach to delivery. It also recommended that increased

evaluation of aid for trade be promoted and funded, with appropriate methods to evaluate aid-for-trade initiatives being developed at programming and policy levels. In particular, the aid-for-trade policy community should answer the question of how to address the specificities of evaluating aid-for-trade activities, as opposed to other development programmes and projects.

There is a growing need to demonstrate that aid-for-trade programmes and projects have been well implemented and have achieved their development objectives. The Aid for Trade Initiative has helped to mobilise more aid for trade. This has increased pressure to show not only that the Initiative is a success in terms of the process it has established, but (more importantly) that it has led to positive changes in partner countries with regard to achieving these countries' trade and development goals. Therefore, much more ought to be known about what works and what does not.

As the donor community has scaled up its aid-far-trade activities, a discussion has begun on how to measure the outcomes and impacts of these activities.[2] The Director-General of the WTO, Pascal Lamy, has outlined future work on evaluation of aid for trade as follows: "*Aid for trade should develop as a community of best practice. A first step in this direction is to inventory what is out there. The second is to look at common frameworks. We need to ensure also that we are not just measuring inputs and outputs, but tangible outcomes."*[3]

1.2 What should be evaluated

An aid-for-trade evaluation requires a common understanding of what aid-for-trade programmes and projects are expected to deliver. In particular, what are the goals, specific objectives and key results? Evaluative research needs to begin by investigating the problem that a policy, programme or project is addressing. Objectives at the programme or project level need to be clearly defined. Data needs and gaps should be determined to enable evidence-based evaluation (Ravallion, 2008). Moreover, the evaluation itself should be based on a firm definition and clear objectives.

The Task Force considered programmes and projects to be aid for trade if they had been identified as trade-related development priorities in recipient countries' national development strategies. At the same time, clear and agreed benchmarks are necessary for reliable global monitoring of aid for trade. The following set of six broad categories summarises the aid-for trade-agenda: (i) trade policy and regulation; (ii) trade development; (iii) trade-related infrastructure; (iv) building productive capacity; (v) trade-related adjustment; and (vi) other trade-related needs. Within each category there can be great diversity in programme or project types (as well as in feasible evaluation approaches and metrics) (OECD, 2007). Some donors adopt a narrow definition of the aid-for-trade agenda (*e.g.* only trade policy and regulation and trade development), while for others it consists of a broader subset of activities identified by the Task Force.

Basing the monitoring of aid for trade on the OECD Creditor Reporting System (CRS)[4] implies that all aid-funded investments in transport, energy and telecommunications are considered to be trade-related. However, such investments have or could have outcomes that are not primarily trade-related. This raises the question of how to evaluate programmes and projects that are not primarily trade-related, but are part of the broader aid-for-trade agenda.

Furthermore, some assistance from multilateral and regional development banks (*i.e.* investments in the tradable sector and trade finance) takes the form of non-concessional lending or low concessional financing, which have a development impact but are not counted as aid for trade (World Bank, 2009). Nevertheless, these Other Official Flows (OOFs),[5] particularly in the area of trade development and infrastructure programmes and projects, do build capacities. This supports the case for incorporating them in aid for trade evaluation.

Some donors channel aid through multilateral agencies. The Multilateral Organisation Performance Assessment Network (MOPAN), a group of 16 donor countries with a common interest in monitoring multilateral effectiveness, appears to be an appropriate forum in which donors can co-operate on joint monitoring and evaluation of aid-far-trade programmes and projects undertaken by multilateral agencies.[6]

1.3 How evaluations should be conducted

There are many different possible types of evaluations, with both qualitative and quantitative dimensions. Every evaluation involves trade-offs. Approaches need to be robust enough to be meaningful, flexible enough to be suitable for local needs, and relatively straightforward to execute. The simplest type of comparison is an examination of an indicator to see how it has changed following an intervention. This single-difference comparison is the approach most commonly used.[7] However, introducing this type of comparison in aid for trade would be extremely difficult, as there is no appropriate control group and interventions can have very different effects depending on their context.

The Millennium Challenge Corporation (MCC) in the United States has adopted the following criteria for choosing evaluations: (i) the need for information; (ii) the potential to learn from an evaluation; and (iii) the cost and feasibility of carrying out impact evaluations (IEs). The MCC's use of IEs[8] is based on the recognition that there have been too few independent and rigorous evaluations (Wiebe, 2008). Ravallion (2008) has outlined a number of steps that could make impact evaluations more relevant, some of which appear especially important in evaluating aid for trade. According to Ravallion, evaluators need to take a comprehensive approach to sources of bias, make greater efforts to measure spill-over effects, fully explore impact heterogeneity, understand what determines impact, and develop capabilities for evaluation within developing countries.

The World Bank Independent Evaluation Group (IEG) has found that those responsible for evaluations conduct a wide range of them, and that many evaluations establish impact deductively.[9] There are a number of constraints on the use of IEs. These constraints affect some donors more than others, depending on their resources and expertise. The MCC has noted that evaluations can be very costly.[10] It has also suggested that the findings of midterm reviews and evaluations could require modifications to the implementation plan that are unacceptable. Moreover, some programmes and projects are so idiosyncratic that conducting an evaluation would not generate useful information. In the case of some aid-for-trade programme and project designs and sectors, it is impossible to identify a reasonable control group.

Impact assessment can be hindered by a lack of clear targets and data. There is currently a significant data gap with regard to impact assessment in many areas, including trade. In evaluating aid for trade, the focus cannot be on micro interventions through aid-far-trade programmes and projects alone. Broader policy considerations are equally

important if not more so.[11] For trade, economy-wide macro reforms and actions rather than micro interventions can be the real drivers of change.

1.4 Results that can be expected

The purpose of evaluation is to increase accountability and to learn from project implementation – establishing what works, under what circumstances and why. Evaluation also helps to understand what has not worked, and where improvements can be made. By holding all parties to account, it provides incentives for donors and partners to deliver on their commitments and ensure that aid is delivered and employed effectively in order to achieve maximum development gains. Beyond providing scrutiny, evaluation should have a learning objective. However, findings are influenced by their context and may not be generalisable. They should be interpreted with care. An important emerging question is how to learn from evaluating aid-for-trade programmes and projects across countries.

As noted, the WTO Task Force outlined multiple goals for aid for trade. They include increasing trade, diversifying exports, maximising linkages with the rest of the economy, increasing adjustment capacities, regional integration, and contributing to inclusive growth and poverty reduction. Nevertheless, constraints may prevent programmes and projects contributing to trade growth. Outcomes are therefore a function of aid-for-trade programmes and projects, but also of complementary policies. "Behind the border" issues can constrain trade growth and diminish its impact on overall economic growth. While measuring the impact of programmes and projects is an important goal, their effectiveness depends critically on the overall investment climate and on political leaders' commitment to promote trade (World Bank, 2009).

Attribution is problematic in many evaluations. For instance, in the case of budget support (particularly at outcome and impact levels) effects are "difficult to disentangle from other influences (including the influences of other aid flows)" (Dom, 2007). Although monitoring inputs and outputs is straightforward, it can be hard to identify outcomes that are definitely attributable to aid for trade. Clemens, Radelet and Bhavnani (2004) have argued that it is inappropriate to conflate all forms of aid and expect an impact on growth over a relatively short period of time. However, they have distinguished three types of aid that might have different correlations with growth:

- emergency and humanitarian aid, which is likely to be negatively correlated with growth in many cases because of the effects of conflict;
- aid that affects growth only over a long period if at all (*e.g.* aid to support democracy, the environment, health or education);
- aid that could plausibly stimulate growth over four years, including budget and balance of payments support, investments in infrastructure, and aid to productive sectors such as agriculture and industry.

They found that aid in the third category was associated with higher growth. Elliot (2007) has made the point that there is considerable overlap between this third category and the broad definition of aid for trade. Aid for trade is different from many other forms of development assistance in that it targets productive (tradable) sectors, whose development can contribute to increased economic growth in the short to medium term

(*i.e.* within a few years). While this may be true of some elements of aid for trade, other elements can have an even longer-term impact. For instance, New Zealand has reported that "we recognise that many of the challenges facing countries to take advantage of trade opportunities will take time to address. We endeavour to take this into account in setting objectives and timeframes for our trade and development policy and programmes."[12]

Given the broad nature of aid for trade, focusing on trade outcomes alone may be too narrow. While some elements of aid for trade can also achieve development objectives (*e.g.* assisting with agriculture and food security or providing the social benefits associated with transport systems), a number of constraints prevent gains in trade from being realised.[13] How outcomes may be linked using indicators is explained in this report. The overview provided is a useful way to start a discussion on the immediate, intermediate and final objectives of using the results of aid for trade to address supply-side constraints.

1.5 Current practices

This section provides an overview of donors' experience in evaluating aid for trade, the guidelines and approaches they use, and different levels of evaluations. It also highlights some of the challenges they face.

Donor experience

The OECD's DAC Evaluation Resource Centre (DEReC) database is an inventory of evaluations of trade-related activities.[14] It allows users to access and learn from a wide array of key evaluation publications, including those related to trade and infrastructure, quickly and easily. A source of knowledge and data much like the CRS, this database has been underutilised by the aid-for-trade community. Discussions will take place at the OECD with the evaluation network to make it more accessible.

Few evaluations specific to aid for trade have been undertaken to date. This is likely due to a lack of sector-specific approaches and procedures, but also to the Aid for Trade Initiative having emerged only recently as a distinct category of development co-operation. However, evaluations of trade-related assistance and other activities have been undertaken in the categories now covered by aid for trade.

Different donors make use of different definitions, objectives, approaches and resources. Evaluations of trade-related assistance typically involve desk studies, field interviews, and a mix of qualitative and quantitative methods. Some donors use specific databases for quantitative analysis, while others rely on questionnaires (OECD, 2006). Most have evaluation guidelines, which they continually revise. For instance, Japan has examined and revised its evaluation guidelines three times. The fifth version is being drafted. Japan regularly reviews and upgrades its capacity and methods for evaluating development co-operation programmes and projects, including in the area of aid for trade.[15]

Many evaluation guidelines are based on the OECD Development Assistance Committee (DAC) *Principles for Evaluation of Development Assistance* (OECD, 1991) (Box 1.1). Responses to the 2008 OECD/WTO Donor Questionnaire on Aid for Trade revealed that 86% of donors monitor and evaluate their aid-for-trade programmes and projects in accordance with generic evaluation guidelines. Few donors or agencies use specific guidelines for themes and sectors that fall under aid for trade. Most evaluate their

aid-for-trade programmes and projects as part of more general efforts to evaluate Official Development Assistance (ODA) through structured and regular evaluation exercises.

Box 1.1 The DAC criteria for evaluating programmes and projects

When evaluating programmes and projects, it is useful to consider the following DAC criteria, as laid out in the DAC *Principles for Evaluation of Development Assistance*:

Relevance: The extent to which the aid activity is suited to the priorities and policies of the target group, recipient and donor. In evaluating the relevance of a programme or a project, it is useful to consider the following questions:

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

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

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

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

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

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

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

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

Sustainability: Sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn. Projects need to be environmentally as well as financially sustainable. When evaluating the sustainability of a programme or a project, it is useful to consider the following questions:

- To what extent did the benefits of a programme or project continue after donor funding ceased?

What were the major factors which influenced the achievement or non-achievement of sustainability of the programme or project?

Source : OECD (1991)

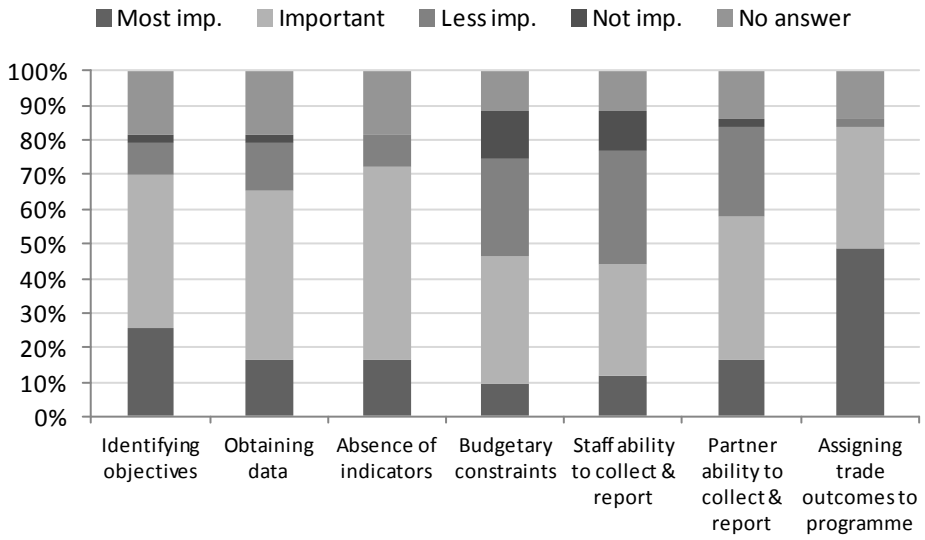
Many donors are strengthening their evaluation frameworks for aid for trade. The European Commission (EC) undertakes evaluations at mid-term and at the end of programmes and projects, where appropriate. The evaluations are "tailored to the context and type of aid-for-trade operation".[16] Finland and Sweden plan to strengthen their monitoring and evaluation systems by developing specific indicators. Germany is looking at ways to design a mechanism, including impact chains and indicators, for monitoring the implementation of the Economic Partnership Agreements and plans to pilot the mechanism in an African, Caribbean and Pacific (ACP) country. Ireland and Spain will review their monitoring and evaluation frameworks in the context of their new strategies (OECD/WTO, 2009).

France, as part of an ongoing evaluation of its trade capacity-building programme, studied ways to improve the quality of different aid-for-trade activities between 2002 and 2008, with a total value of USD 60 million. An independent evaluation in 2009 highlighted the value added of the French aid-for-trade programme, while pointing out that measuring the impact of such programmes on international trade was impossible due to lack of data and attribution problems (Agence Française de Développement, 2009).

USAID recently concluded an independent evaluation of its trade capacity building programmes and projects, with a total value of USD 1.5 billion, implemented between 2002 and 2006. The single most significant problem encountered by evaluators was obtaining programme- or project-related documentation. A secondary problem was the absence of independent evaluation documentation. Nevertheless, this evaluation established that trade capacity building programmes and projects have a positive effect on developing countries' exports, as well as discernable employment and income impacts on individuals and families. It also found that these programmes and projects did not appear to impact export diversification or export concentration at the national level. It noted that world demand and prices, growth rates, and domestic economic and business policies are more important factors influencing export performance.

The 2010 OECD/WTO donor questionnaire included questions on the problems donors face in evaluating aid for trade. Some have budgetary constraints and have experienced difficultly in collecting data (both aggregate data and data whose source is partner countries). Others cited a lack of indicators for tracking progress. The problem identified as the most challenging was how to assign trade outcomes to aid-for-trade programmes and projects (Figure 1.1). Programme monitoring by the EC focuses on output and possibly on outcomes, but it "does not consider it realistic to monitor trade impacts of specific aid programmes because of the important number of external factors influencing trade." The United States pointed out that because "success in trade can involve so many variables... it is very difficult to monitor results in this area and attribute results to specific programmes." Its discussions have focused on sharing lessons learned and best practices, which it considers the most productive approach. On the other hand, the World Bank considers that this field still relies excessively on outdated methods compared to other fields of development work: "An energetic push for the adoption of impact evaluation techniques and their mainstreaming in project design is needed." The remainder of this chapter examines potential techniques for overcoming constraints on evaluating aid for trade.

Figure 1.1 What donors find most challenging in evaluating aid for trade



Source: OECD/WTO (2011)

Evaluators generally find it difficult to measure impact, especially in the long term. Although trade-related support appears to be well implemented with regard to delivering inputs and producing planned outputs, it has not been possible to draw firm conclusions on outcomes and impacts further along the results chain. This does not necessarily mean that programmes and projects have been unsuccessful. The problem is related more to the methods used. At programme or project level, it is usually not possible to attribute higher level outcomes and impacts to specific contributions. Evaluators could argue that trying to do so would not be a meaningful exercise in most cases.

Levels of evaluation

Evaluations can be carried out on many different levels. For instance, Japan conducts them at policy, programme, sectoral and project levels, using various methods and targeting a wide range of ODA issues including aid for trade. Aid-for-trade outcomes can be considered at the country and regional levels. The EC carries out specific evaluations of co-operation by country and by region, and there is a specific focus in regional evaluations on regional integration.[17] In the United Kingdom, the Department for International Development (DFID) is examining ways to evaluate its overall aid-for-trade strategy.[18]

Country studies are joint evaluations of all trade-related assistance programmes and projects in a country, with attempts to determine whether joint efforts are leading to impacts at the macro level. The United Nations Evaluation Group (UNEG) is currently evaluating the impact of country level joint programming in eight pilot countries under the UN Chief Executives Board (CEB) Inter-agency Cluster on Trade and Productive Capacity (OECD/WTO, 2009).

In addition, portfolio and product level evaluations have yielded useful information:

- *Portfolio level:* The World Bank Independent Evaluation Group (2006) evaluated the Bank's support for trade, which amounted to USD 38 billion in 117 countries over the period 1987-2004. The IEG found that, on the whole, the Bank was too optimistic about the ultimate benefits of trade: "The Bank was effective in helping developing countries liberalise trade regimes, but trade initiatives were less successful in generating a dynamic and sustained export growth and diversification path." It also found that the Bank had underestimated the need for complementary policies. The report recommended that trade reform be better combined with complementary policies, investments and institution-building measures.

- *Product level:* Brenton and von Uexküll (2009) have looked at the impact of product-specific export development programmes and projects. For those targeted at a specific product, there is a measurable outcome variable: export of that product. Based on the examination of a number of cases, the authors concluded that exports increase as a result of export development programmes and projects.

The overview of donors' experience shows that they use varying definitions, objectives, approaches and resources. Thus, neither common evaluation frameworks nor new guidelines are required for aid for trade. Generic approaches are considered sufficient. The evaluation community has seen guidelines specially tailored for different areas proliferate, with little added value. On the other hand, a discussion concerning conceptual challenges, particularly linkages in results chains, the attribution problem and the sequencing of various programmes and projects, could bring about greater operational clarity.

1.6 Approaches and methods

The weak trade performance of low-income countries is due in part to poor trade and transport infrastructure. Lack of transport infrastructure is a key binding constraint that aid for trade aims to address. Djankov, Freund and Pham (2008) have demonstrated that trade delays reduce exports, and that the effect is greater than that of tariffs. ODA flows in 2008 showed that 37% of sector allocable ODA went to aid-for-trade programmes and projects, and that overall aid for trade had increased to USD 39 billion (2008 constant). There was a large increase during this period in aid for transport and storage in general, and road transport in particular. Evaluation in this area is becoming increasingly important. Consequently, the OECD conducted a meta-evaluation for the transport and storage sectors (Chapters 2 and 3, Annex A) and many of the examples cited in this chapter are related to the evaluation of road upgrade programmes and projects.

Results chains

Because of the "black box between outputs and impact",[19] a results chain needs to be clearly articulated and attempts should be made to identify specific impacts and how they are linked. Programmes and projects should have specific objectives, and there should be a clear idea of how these objectives will be achieved. Potential results need to be considered. For instance, upgrading and expanding a roadway network (primary, secondary or rural) can reduce transportation costs, improve the level of service, and provide important linkages to national, regional and international markets. Rural roads can provide the poor with access to social services, including health and education, as

well as markets. Given the importance of investments in road upgrades within aid for trade, rigorous impact evaluations can provide lessons for future funding.

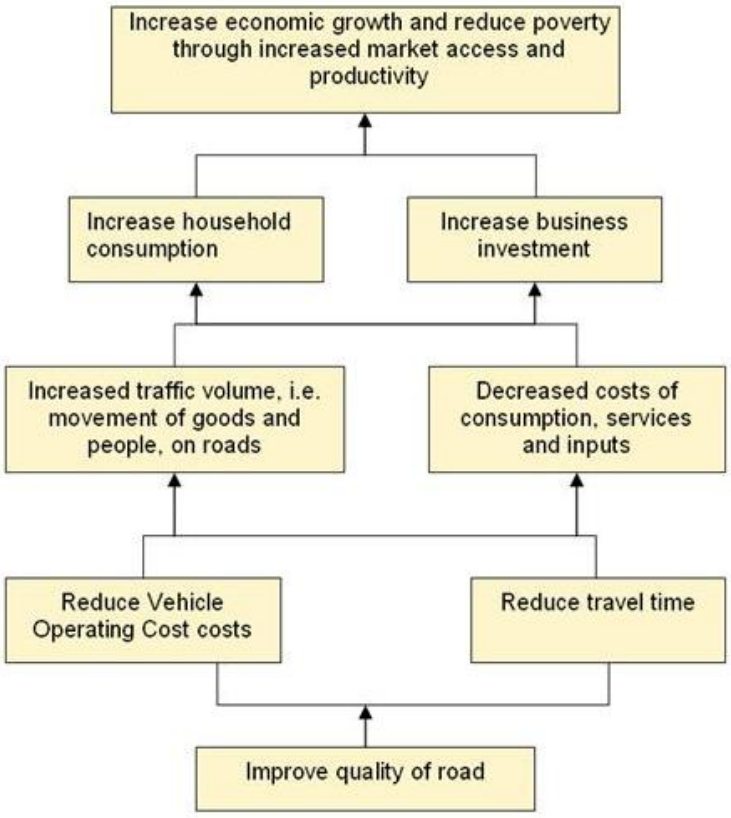
Cost-benefit analysis can be useful in establishing criteria for undertaking specific programmes and projects, *i.e.* identifying particular road upgrade activities. Such analyses typically quantify two main benefit streams:

- saved travel time due to the ability to drive faster on better roads, yielding, in turn, increases in household income through reduced prices for goods and services and increased business investment owing to reduced costs for inputs and firm value added;

- reduced wear and tear on vehicles, yielding savings in vehicle operation and maintenance expenditures.

A logic model (Figure 1.2) can be used to identify expected outcomes. The model shows links in the chain, from the outcome (improved road quality) to national level impact (economic growth and poverty reduction). The identification of outcomes is essential in justifying particular programmes or projects, but also in quantifying the expected results, which can then be used to track progress and evaluate whether the ultimate objectives have been met.

Figure 1.2 Millennium Challenge Corporation logic model for improved road quality

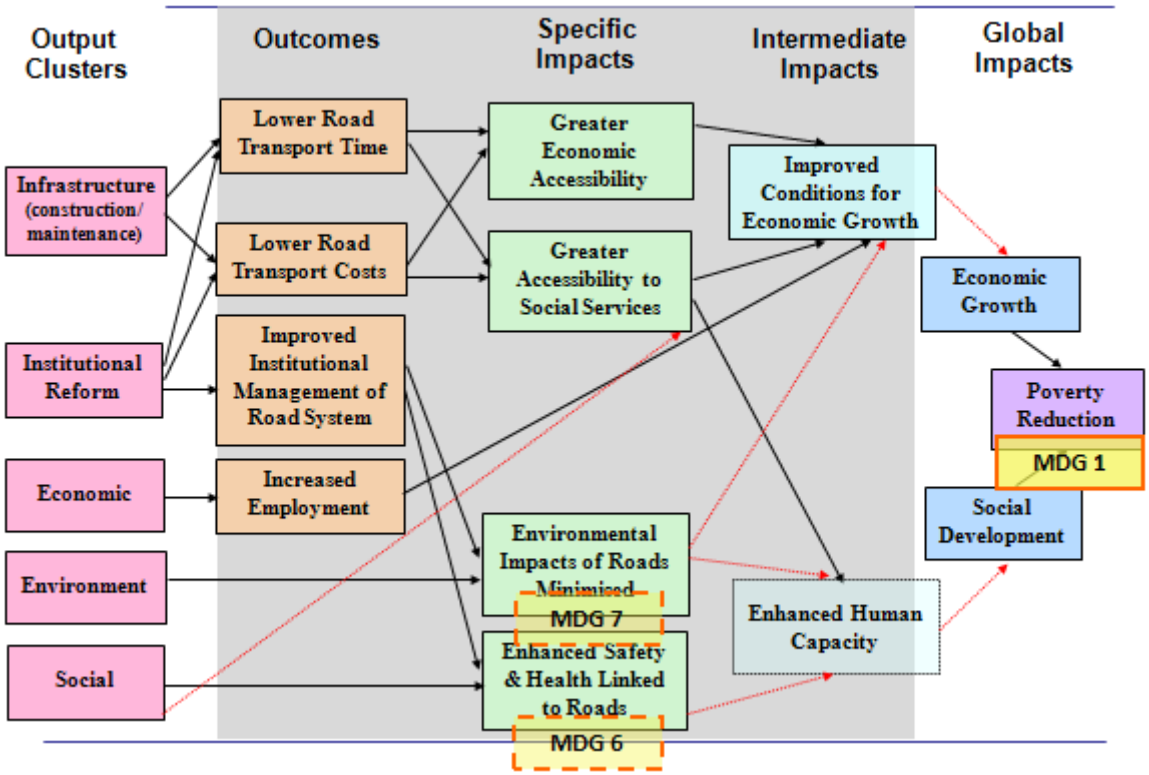


Source: "Millennium Challenge Corporation, Impact Evaluations, Overview" (*www.mcc.gov/pages/sectors/impact/roads*)

The MCC model considers direct economic impacts and how they are linked, from the output of improving road quality to the overall impact of increasing economic growth and reducing poverty through increased market access and productivity. Impacts in between are largely measurable (*e.g.* were costs reduced and did traffic volume increase?), providing metrics against which to measure success or failure. The logic model does not articulate increasing trade as a direct outcome. However, there is a reference to the increased movement of goods, which indicates commerce if not necessarily international trade.

The EU has developed intervention logic models for improving road quality which are more complicated (Figure 1.3). They highlight economic, social and environmental impacts along the results chain and consider costs and time savings, as well as the impacts of institutional reform, environmental consequences, and enhanced safety and health associated with road upgrades. They also link specific and intermediate impacts to global impacts and to the contribution roads can make to economic growth and social development.

Figure 1.3 EU intervention logic model for improved road quality



MDG: Millennium Development Goal

Source: European Commission *http://www.oecd.org/dataoecd/7/56/46303700.pdf*.

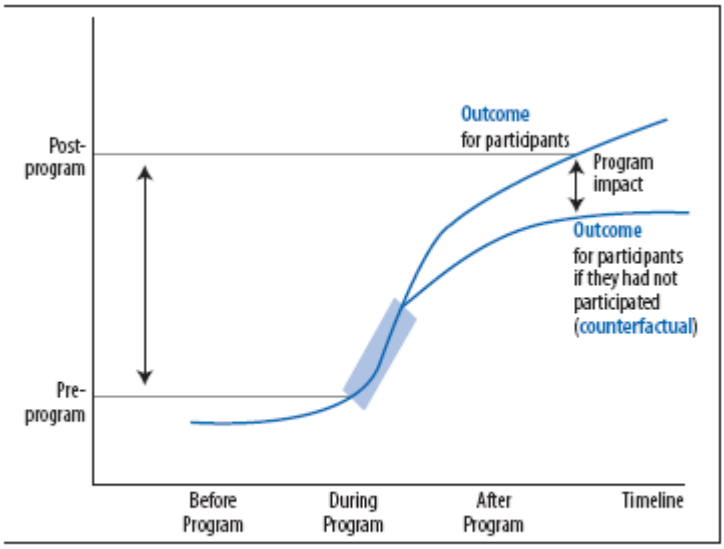
The logic models for improved road quality in Figures 1.2 and 1.3 are potentially useful in considering other aid-for-trade interventions. However, they are simplified tools, not templates. An intervention logic needs to be developed for each country, based on its context and existing government strategies. At country level, logic models are likely to be

similar. The development of general models to stimulate thinking about the series of impacts of aid-for-trade interventions would be helpful with regard to establishing evaluation questions and the data needs of an evaluation. It could also be beneficial in providing specific logic models/intervention logics for aid-for-trade that focus on intermediate and final trade impacts. While the models clarify expected outcomes and impacts, the impacts are not fully attributable to development interventions. Outputs are 100% attributable to these interventions, but for impacts the certainty of attribution decreases as one moves along the chain of results.

Impact evaluation

The growing political demand by academics, NGOs, international agencies and voters for results in development has stimulated new thinking about how such results may be demonstrated. Impact evaluation is one approach currently being applied in the evaluation of a range of development activities.[20] It is an approach that focuses on "with and without" interventions, and compares outcomes for participants in a programme or project with a control group. Figure 1.4 represents the timing of a programme or project and how the impact is determined by comparing outcomes for participants against the counterfactual. Impact evaluation typically involves a large number of units of intervention, *e.g.* children, households, firms and schools. Examples of evaluations involving a small number of observations are those of policy reform and many (but not all) capacity building programmes and projects. These are more amenable to qualitative approaches.

Figure 1.4. Programme or project timing and impact determination



Note: The impact of a program is the difference in the outcome variable for participants and what participants would have obtained if they had not participated.

Source: Fitzsimons and Vera-Hernandez (2009).

In the evaluation of aid-for-trade programmes and projects, identifying appropriate control groups can be difficult. Moreover, the application of impact evaluation to aid for

trade is hindered when targets and data are unclear. Without baseline data and a controlled experiment, it is difficult to attribute a programme or project's success or failure to the programme or project itself, as opposed to the environment in which it operates and the unpredictable shocks (positive or negative) that influence its effectiveness (Birdsall, 2004). An additional problem is that programmes, projects or policy reforms do not take place in isolation. They may interact, making identification of the actual benefits precarious. Nevertheless, randomised trials have potential: they "revolutionised medicine in the twentieth century, and have the potential to revolutionise social policy during the twenty-first" (Duflo and Kremer, 2008).

The World Bank and others are investigating how to improve trade-related impact evaluations. The Inter-American Development Bank (IDB) has carried out a number of *ex post* evaluations of export promotion. The World Bank is working on a similar exercise for Tunisia. The Bank is likely to pay special attention to trade facilitation. In 2010, it organized a workshop that brought together researchers to discuss randomised control trials and their application to trade.[21] Work in the area of trade promotion is promising due to the more disaggregated nature of treatment at the firm level. However, since a number of aid-for-trade programmes and projects finance public goods and affect macro policies, it is unlikely that this tool can be widely applied to aid for trade. The MCC has applied randomised control trials to rural road upgrades. While the experiment did not succeed, it provides insights into how such studies could be undertaken, including data needs and methods.

The Millennium Challenge Corporation's rural road upgrade project

Randomised impact evaluations are normally conducted at the individual level (*e.g.* in education or public health). Establishing a credible comparison group for an aid-for-trade intervention would therefore be difficult, but not impossible.

The six rural road improvement evaluations currently being undertaken by the MCC attempt to compare similar roads, some of which have been upgraded and others not. The roads which have not been improved essentially act as a control group.[22] Road upgrades in some areas may take longer to achieve. For instance, they may not constitute the most important bottleneck in the region or may otherwise be a low priority.

Improved transportation infrastructure gives rise to complex economic interactions. Determining impacts at different levels, as outlined in the MCC logic model (Figure 1.2), can be hampered by lack of data, lack of a historical track record, and difficulties in measuring infrastructure services. Data collection is often the most costly element of evaluations. Once staff are trained, cross-cutting a sample of projects can dramatically reduce costs. Household surveys are an important way to find out whether a programme or project has had the desired effect on the most appropriate group to use in determining its success, *i.e.* road users. Surveys can help establish an appropriate baseline from which changes may be measured. Follow-up surveys can help determine the changes that have resulted from road improvements.

Surveys can be used to look specifically at impacts or changes at the household and firm level, taking account of local prices and incomes in the area and how they evolve. For instance, what were the growth rates of these variables before and after the programme or project was implemented?

This evaluation design considers local trade impacts. Impacts on international trade would be more complicated to determine due to a number of spill-over effects, but many

of the expected impacts are testable. For instance, did the existence of a road lower input prices? Did farmers switch to export crops?

Wider applicability

This approach may work with regard to the upgrading of rural roads, but what about large national level programmes or projects such as those involving ports or airports? In these cases, interim indicators can be used (for further discussion of indicators, see Chapter 4). As an example, the port of Cotonou in Benin is a major hub with no comparable port facilities, making it impossible to carry out impact evaluation.

Nevertheless, there are ways to infer impacts. To do so, considerable work needs to be done in advance of the programme or project to determine the changes that would be expected in terms of the output prices and quantities of specific products traded. The MCC first undertakes *ex ante* analysis (*e.g*. due diligence, cost-benefit analysis) targeted at price and income impacts at household level and specific products and sectors. For instance, did prices and incomes change more rapidly after the programme or project was implemented? The MCC examines its logic model to determine whether it is consistent with evidence on the ground. Did the programme or project differentially improve interim indicators?

It is important to determine impacts at the micro level so as to establish the aggregate effect on the wider economy. Macro level impacts can be determined through modelling. A number of assumptions need to be made based on the changes that evaluators hope to see (*e.g*. in the time it takes to clear customs). Parameters from the evaluation can be imbedded in computable general equilibrium (CGE), input-output, gravity, and other types of models to arrive at simulated changes at the macro level.

Although the methods described above have limitations, they represent promising approaches. The potential problems are that assumptions may drive the results, and that developing a CGE model is extremely costly and time consuming, while existing models such as the Global Trade Analysis Project (GTAP) and the General Algebraic Modelling System (GAMS) may not be suitable for a specific programme or project.

1.7 How should the evaluation process change?

While the importance of evaluation is emphasised here, it is important that evaluation be viewed in an appropriate context. It can help provide answers, but these answers depend on the methods, tools and assumptions associated with the counterfactual used. More constraining than the constraints related to the methods used are those associated with political economy. Evaluations should highlight failures as well as successes, but this is rarely in the interest of donors. Negative evaluations could leave their work open to political criticism and limit their access to new resources. It may therefore be more convenient and less risky politically for them to minimise serious evaluation work. Under these conditions, there are also few incentives to work with other donors or partner country representatives.

This section suggests that evaluations should be collaborative and not be constrained politically with regard to what they can tell us. That is, they should be independent and carried out jointly, where possible in a way that involves partner countries.

Evaluations should be independent

There are currently few independent evaluations of aid. Many donors have independent units for conducting evaluations, but their independence depends on the structure of the institution. Nancy Birdsall, President of the Center for Global Development, has called for the creation of an independent entity to undertake aid evaluations, which could easily be funded by small proportional contributions from each donor. She has suggested that the critical ingredients of an evaluation of development assistance are that it should be independent, collectively agreed, and financed by a minimum set of large multilateral and bilateral donors (Birdsall, 2004).

We should not expect to be able to answer every question that could be asked at the political level. However, political expediency may influence the choice of evaluation procedures. While truly independent evaluation can highlight successes, "as chance might have it, the evidence may just refuse to co-operate. Hard evidence is simply not worth the trouble, especially if eloquence and a few carefully chosen examples can carry the day" (Banerjee and He, 2008).

Evaluations are usually launched by advocates of certain sectors who want to showcase their achievements. In fact, true believers see no intrinsic value in rigorously testing the policies they advocate. There is strong support for trade leading to development results, but more needs to be understood about when and where aid-for-trade interventions make this possible.

Independent evaluations of trade-related assistance would be a valuable global public good. Through these evaluations, it would be possible to know much more about what truly works and, crucially, what does not. It would also be possible to learn more easily about how to improve practice, and results could be widely distributed. In addition, economies of scale could exist (with significant cost savings) since donor agencies would not all need their own teams of evaluators.

Evaluations are expensive. Full use should be made of them, and each new evaluation should contribute to the ongoing debate. There is currently little discussion of evaluations at the global level, where overall lessons should be discussed. An independent source of evaluations would help to facilitate this discussion.

Joint evaluations should be encouraged

Teaming up with other donors to conduct evaluations can be demanding. It is often easier to work alone. Many donors want to know the specific contribution their taxpayer money is making to the outcomes of a development assistance initiative. While joint evaluations make sense from a learning point of view, they are not always compatible with accountability.

Joint evaluations seem particularly appropriate in the case of aid for trade, where efforts are made to establish the outcomes and impacts of a range of trade-related programmes and projects on overall competitiveness and trade performance. Good knowledge is needed about the disaggregated effect of different interventions. For instance, determining the impacts of a port upgrade may require some time, especially if some of these impacts depend on other interventions such as upgrading farm-to-port roads. If the facilitation of customs procedures and trade is also considered, what are the additional effects?

One of the innovative aspects of the Aid for Trade Initiative is its focus on sequencing a number of interventions from different donors (*i.e.* trade policy advice, trade development and trade-related infrastructure) into a coherent framework for building trade capacities. Rather than undertaking all these interventions simultaneously, co-ordinated sequencing of programmes and projects would enhance the overall outcome. In order to learn more about the impacts of individual interventions, and overall impacts when all parts are completed, donor co-ordination is needed.

Partner countries should be involved in evaluations

"Is this information you are gathering from us just to help you write your report or can you really be helpful to us?"

– a woman in South Sudan (Prichett, 2009).

The aim of evaluation is ultimately to improve the effectiveness of development interventions. Successful evaluation cannot be a passive activity. Neither can it just be about gathering information to write reports. Soliciting partner country input in designing an evaluation is crucial if useful information is to be obtained. Nevertheless, practical constraints can prevent effective partner country involvement.

Donors often conduct evaluations with partner countries. Some donors have assisted partner country governments in carrying out programme or project evaluations. A number of multilateral and regional institutions complete programme or project reports in consultation with partner country stakeholders (OECD, 2009). Partner countries participate in the WTO and should be involved in discussions about evaluating aid for trade so that their views are heard.

1.8 Options for better assessment of aid-for-trade programmes and projects

A key lesson to be drawn from this report is that there is no common pattern in the way donors evaluate their programmes and projects. This implies that current guidelines may not be precise enough, or that they may not be implemented strictly enough. A review of existing guidelines is beyond the scope of this report. It seems more useful to provide some information on economic techniques used relatively frequently in other areas (*e.g.* to assess the impact of future regulatory changes) in order to take into account the essential *ex ante* aspect. In this context, Table 1.1 presents in simplified form the series of questions the Australian Office of Best Practice Regulations has instructed evaluators to ask when the potential impact of regulatory changes is assessed.

Donors and recipients could adapt such a list to the aid-for-trade agenda. It would not replace interviews or field information. Ideally, it would be combined with feedback from those directly involved in aid-funded operations. The use of such a list would not eliminate the need for constant review, which serves both to assess the sustainability of programmes and projects over time and to train evaluators.

Table 1.1 Impact assessment checklist adaptable to evaluations of aid for trade

Step	**1.**	**Identifying the problem**
	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
Step	**2.**	**Defining the objectives of government action**
	2.1	Defining the objective
	2.2	Defining the outcome
Step	**3.**	**Examining the options that may achieve the objective/outcome**
	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?
Step	**4.**	**Impact analysis – costs, benefits and risks**
	4.1	Who is affected by the problem?
	4.2	Who is likely to be affected by the proposed solutions?
	4.3	Costs
		To producers
		To consumers
		To the community and/or environment
		To governments
	4.4	Benefits
		To producers
		To consumers
		To the community and/or environment
		To governments
	4.5	Analysing risk
	4.5	A few key additional points
		Competition assessment
		Effect on small businesses
		Effect on trade
		Ecologically sustainable development
		Deviation from international standards
	4.6	Quantifying the impacts where significant
		Valuing costs and benefits where there is no market
		Discounting
		Sensitive analysis
		Quantifying the compliance costs
Step	**5.**	**Consultation**
Step	**6.**	**Conclusion and recommended option**
Step	**7.**	**Implementation**
Step	**8.**	**Review**

Source: Adapted from OBPR (2010)

With these qualifications, several specific points need to be made about the checklist. First, *ex ante* Steps 1 to 3 are essential and should be well developed. Step 1 entails 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. These, in turn, require a thorough analysis of the relevant regulations in place. Step 2 is about clarifying and defining objectives and ranking them according to primary and subsidiary outcomes. Too often, an operation is treated as a "stone" that can "kill two birds". Step 3 is about examining all available

options, from no action, to market-based instruments (*e.g.* tradable property rights, auctions), to subsidies, to taxes, to regulatory reforms, and selecting the appropriate response. These options are defined broadly, so that international aid is included.

Step 4, which focuses on impact analysis, has two aspects. First, it should be carried out 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 programme or project. For instance, costs should include administrative costs, changes in sources of supply, and changes in input prices. Costs to consumers should include not only changes in the price of a good or service, but also reductions in "utility" (*e.g.* quality, range of choice), undesirable income or wealth redistribution effects and unemployment, among others. Benefits to consumers should be examined just as comprehensively.

Steps 5 to 8 deal with the inter-active process, *i.e.* between policy makers and the people directly affected by a programme or project. Here, too, the importance of the review process (Step 8) cannot be over-emphasised. If an evaluation is conducted fairly, it can help to build trust between policy makers and others. Trust may well be the most crucial input for efficient aid for trade.

1.9 Discovering what works at the global level

In evaluating aid for trade, the focus should not only be on micro interventions and aid-for-trade programmes and projects. Broader policy considerations, which are equally if not more important, also need to be considered. With regard to trade, economy-wide macro reforms and actions rather than micro interventions can be the real drivers of change. Thus, it is a necessary to be able to bring evaluation findings together to infer broader lessons.

Good programme or project level work is necessary before looking at macro/country level aggregates. In addition, there is a political need to consider the macro impacts of aid for trade. Discussions at programme or project level must therefore feed into broader discussions at the country, regional and global level. The OECD and the WTO can help provide a global perspective, bringing evaluations and other sources of information together and drawing conclusions about broader lessons and impacts. The tools described below seem particularly appropriate to provide this perspective:

Meta-evaluations

Meta-evaluations (evaluations of evaluations) are designed to aggregate findings from a series of evaluations. For instance, they make it possible to gauge the extent to which trade is considered in evaluations in areas related to aid for trade such as transport, energy, agriculture and forestry. A meta-evaluation can determine impacts at country level and provide better guidance on programme or project implementation and design, especially how to take better account of trade objectives in trade-related development programmes and projects. To that end, the OECD recently carried out a meta-evaluation focusing on evaluations in Vietnam and Ghana and in the transport and storage sectors (Chapters 2 and 3 and Annex A).

Synthetic reviews

Synthetic evaluations are similar to meta-evaluations, but they are more comprehensive in their scope and use of sources. The International Initiative for Impact Evaluation has launched a series of these reviews to assess the impact of development interventions. They go further than a standard literature review because they aim to provide an unbiased assessment of what works and why through a systematic identification of relevant studies and a synthesis of quantitative and qualitative evidence.[23] The purpose of a systematic review is to sum up the best available research on a specific question.

A systematic review uses transparent procedures to find, evaluate and synthesise the results of relevant research. Procedures are explicitly defined in advance, in order to ensure that the exercise is transparent and can be replicated. This practice is also designed to minimise bias. Studies included in a review are screened for quality, so that the findings of a large number of studies can be combined. Peer review is a key part of the process. Qualified independent researchers control the authors' methods and results. The constraint with regard to aid for trade may be the number of known studies. In addition, very few aid-for-trade actors would have the resources to conduct these reviews. A synthetic review involving the compilation of over 1 900 papers, from which over 50 have been chosen for careful study, could be considered typical.[24]

Econometric analysis

A number of analytical research streams are relevant to the evaluation of aid for trade. Such work can provide a broader focus on aid for trade and its impacts beyond the programme or project level. For instance, cross-country statistical analysis can examine patterns of aid, trade, growth and policy settings and how these influence the overall impact of an aid-for-trade programme or project. Cross-country research can potentially provide information to the trade and development community in a number of ways (OECD, 2007). However, there is often distrust concerning regression findings, particularly in the development community, whereas the trade community is more accustomed to using such methods. On the other hand, case study approaches favoured by the development community tend to be dismissed in the trade community because the findings are not usually generalisable.

How is it possible to determine which interventions work best? Helble, Mann and Wilson (2009) have found that a 1% increase in aid-for-trade facilitation leads to USD 818 million in trade. They have also presented a so-called "rate of return". But even if trade facilitation worked well in the data sample used by the authors in this analysis, can it be expected to work as well with another group of countries (possibly at a lower stage of development)? There may be a danger of overestimating what trade facilitation can deliver. Cali and te Velde (2009) have outlined an approach to assessing the effectiveness of aid for trade at the aggregate level, but there are significant problems with this paper, particularly with the treatment of disbursements and the appropriate lags applied. In fact, quality of aid-for-trade data is a major constraint on their paper. Use of the WTO/OECD Trade Capacity Building database is inappropriate because of the lack of consistency in reporting. This source has also been discontinued. Commitments data from the CRS are probably of the highest quality, but their use makes little sense in assessing aid-for-trade outcomes. Using data on disbursements makes the most sense, but prior to 2006 these

data are of insufficient quality to use in analytical studies and would introduce significant bias to the econometric results.

Nevertheless, this literature appears likely to expand as policy makers try to obtain an aggregate picture of the effectiveness of aid for trade. Such results should be interpreted for what they are. Needs and priorities are not determined in the aggregate, but at country level. These findings might be introduced in discussions with partner countries, but they tell little about individual cases. Does using the aggregate approach make sense in development co-operation? Is it possible to tell? And if each country has different binding constraints, does it matter whether using the aggregate approach makes sense or not? Considerable care must be taken in interpreting and presenting econometric results. While econometric work can produce useful analyses, it should be conducted independently and, where possible, peer reviewed before it is seriously considered in making policy decisions. Otherwise, it risks being used as a tool to promote certain parts of the aid-for-trade agenda.

Applicability

Many in the development community consider that basing policies on hard evidence is impractical. They fear that requiring every initiative to be justified in this way will bias decisions in favour of what is measurable and easy to evaluate. Furthermore, evaluation does not always produce results that justify actions: "relying exclusively on this evidence is tantamount to considered inaction" (Banerjee and He, 2008). A lack of evaluations, and conceptual difficulties in evaluating trade outcomes, may make interventions in other areas such as health more easily justifiable by donors if there is much more reliable (reassuring) evidence for those interventions.

Before the impacts of aid for trade on trade and economic growth can be determined, there is a need to get micro interventions right. It is necessary to gather the right data, adapt evaluation practices, and work together with donors and partner countries.

Time lags are also important. Many programmes and projects require time to show results. While pressure exists to demonstrate impacts right now, this can only be done through collaborative efforts, bringing together many sources of information to obtain a sense of results at the global level.

The issues discussed concerning methods raise questions about applicability. Impact evaluations should only be used where appropriate on pilot projects, or on valuable, representative or important interventions. They are potentially a useful tool to help fill gaps in existing knowledge, but they are only one such tool and should be used in conjunction with others.

Given that the methods discussed above can be contrived or produce conflicting results, how can policy makers be persuaded to make decisions about resource allocations based on their findings? Every policy action does not need to be justified by hard evidence. Macro policy, for instance, is very difficult to evaluate properly. Nevertheless, evidence from a limited number of micro evaluations can provide insights into macro interventions. What is certain is that much more needs to be known. Evaluations currently do not provide the sort of information needed.

1.10 Conclusions

While donors have been evaluating trade-related assistance for many years, it is difficult to measure the outcomes and impacts of aid for trade. A number of conceptual difficulties need to be acknowledged and, where possible, overcome. A range of existing techniques and guidelines can be used, and many of the building blocks are already in place.

Careful planning of evaluations is crucial and should be undertaken from the beginning of a programme or project. A combination of *ex ante* reviews to determine what the programme or project hopes to achieve, and *ex post* assessments to determine what works and what does not, would be useful in advancing work on evaluation.

Aid-for-trade projects can be disaggregated, and emerging methods for impact evaluation have potentially useful applications. Using macro models to simulate macro impacts is also useful, although it will not be an appropriate way forward for all donors given the constraints on their resources and expertise. Joint evaluation work should be explored in this case, both for the pooling of expertise and for realising significant cost savings. Econometric approaches are useful, but early examples from the literature on their application to aid for trade show that there are many interpretations, data and methodological limitations, and lack of consistency in the main findings.

It is as important to raise awareness of the trade dimension in the evaluation community as it is to encourage the trade community to consider more evaluations and become familiar with evaluation approaches, objectives and tools. No tool is perfect, and the application of tools such as impact evaluation provides particular challenges. However, such innovations could also provide essential insights at the micro level. Based on these insights, it will be easier to judge whether more and better aid for trade leads to greater capacity to trade, increased growth, and the achievement of development objectives.

Evaluation should be the topic of ongoing discussion. We will always be looking for ways to do things better. Donors can learn a great deal from each other, and the OECD will provide regular updates on new resources and relevant papers. It is likely that a mix of approaches will be required in the evaluation of aid for trade. Not all forms of aid are readily amenable to rigorous forms of evaluation (OECD, 2007). While demonstrating the quantitative impact is crucial in order to justify the nature and type of future aid-for-trade programmes and projects, some of the expected outcomes are qualitative (and critical to achieving and promoting stated objectives). These outcomes include national dialogue processes, forms of institutional capacity building, simplification of national regulatory frameworks, and reform of national policy settings (*e.g.* with regard to exchange rates). Furthermore, the greatest need seems to be for better use of existing generic and development guidance on evaluation.

Against that background, it is recommended to provide evaluative guidance to policy and management staff for the formulation and implementation of aid-for-trade programmes. Moreover, a number of donors are pushing forward on various aspects of this agenda. Work is being carried out on aid-for-trade evaluation frameworks (DFID; the Islamic Development Bank, IDB), performance assessment frameworks (OECD) and impact assessment (MCC, World Bank)

Evaluations should be carried out collaboratively. Joint evaluations make particular sense in aid for trade and attempts to evaluate the impact of many trade-related projects

on overall competitiveness and trade performance. Ways need to be found to involve partners effectively in evaluation. Some donors have assisted partner country governments in carrying out evaluations of programmes and projects. A number of multilateral and regional institutions complete programme or project reports in consultation with partner country stakeholders (OECD/WTO, 2009).

This chapter has presented a comprehensive overview of how to evaluate aid for trade. The next two chapters examine the recent OECD meta-evaluation of trade-related evaluations. Analysis of the meta-evaluation supports future work on evaluation by looking at approaches to evaluating programmes and projects in detail, as well as providing key insights into what evaluations related to aid for trade currently tell us, and how they could take better account of the trade dimension.

Notes

[2] According to DAC definitions, outcomes are "the likely or achieved short-term and medium-term effects of an intervention's outputs" and impacts are "the positive and negative primary and secondary long-term effects – intended and unintended – produced directly or indirectly by a development intervention". See Figure 1 in OECD, 2009 (reproduced as Figure 4.1 in Chapter 4 of this report).

[3] "We must keep the foot on the gas on Aid for Trade." Press release, Second Global Review of Aid for Trade, 6-7 July 2009, *www.wto.org/english/news_e/news09_e/aid_07jul09_e.htm* .

[4] See "Aid for trade statistical queries" (*www.oecd.org/document/21/0,3746,en_2649_34665_43230357_1_1_1_1,00.html*) and "OECD StatExtracts: Creditor Reporting System" (*http://stats.oecd.org/Index.aspx?DataSetCode=CRSNEW*).

[5] Other Official Flows (OOFs) are defined in the OECD Glossary of Statistical Terms as "Transactions by the official sector with countries on the List of Aid Recipients which do not meet the conditions for eligibility as Official Development Assistance or Official Aid, either because they are not primarily aimed at development, or because they have a Grant Element of less than 25 per cent" (*http://stats.oecd.org/glossary/detail.asp?ID=1954*).

[6] "Introducing MOPAN" (*www.mopanonline.org*).

[7] Presentation at the OECD by Howard White, Independent Evaluation Group, World Bank: "Current Practices on Impact Evaluations", 15 November 2006 (*www.oecd.org/dataoecd/27/54/37634269.ppt*).

[8] The focus of impact evaluations is on counterfactual analysis and outcomes.

[9] Presentation at the OECD by Howard White, Independent Evaluation Group, World Bank: "Current Practices on Impact Evaluations" 15 November 2006 (*www.oecd.org/dataoecd/27/54/37634269.ppt*).

[10] The MCC cites a rigorous study of a five-year agricultural training programme, which might cost several million dollars that could otherwise be used to train more farmers (Lucas, 2011).

[11] See remarks by Bernard Hoekman at the International Lawyers and Economists Against Poverty (ILEAP) meeting on "Aid for Trade and Development: Moving Towards Implementation," University College, Oxford, United Kingdom, 11 April 2008 (*www.ileap-jeicp.org/downloads/oxford_08/meeting-report.pdf*).

[12] New Zealand's response to the OECD/WTO Donor Questionnaire on Aid for Trade, 2008 (*www.oecd.org/dataoecd/43/30/43150000.pdf*)

[13] For a more in-depth discussion of this issue, see J.-J. Hallaert and L. Munro (2009), "Binding Constraints to Trade Expansion: Aid for Trade Objectives and Diagnostic Tools", *OECD*

Trade Policy Working Paper, No. 94, OECD, Paris
www.oecdilibrary.org/oecd/content/workingpaper/5kmlbl6glf5d-en

[14] See: *www.oecd.org/dac/evaluationnetwork/derec*.

[15] Japan's response to the OECD/WTO Donor Questionnaire on Aid for Trade, 2008
www.oecd.org/dataoecd/43/43/43149672.pdf

[16] The European Community's response to the OECD/WTO Donor Questionnaire on Aid for Trade, 2008.

[17] The European Commission's response to the OECD/WTO Donor Questionnaire on Aid for Trade, 2008.

[18] To ensure effective implementation of its new aid-for-trade strategy, the United Kingdom monitors progress on an annual basis. Evaluations were also to be conducted at mid-term (2010) and at final year (2013).

[19] See the presentation at the OECD by Claes Lindahl, Swedish International Development Cooperation Agency (Sida): "Evaluation of Sida's Trade Related Assistance: A desk review of 20 programmes and projects 1995-2009" (*www.oecd.org/dataoecd/57/33/44082931.pdf*), 10 November 2009.

[20] The International Initiative for Impact Evaluation (*www.3ieimpact.org/database_of_impact_evaluations.html*) provides a database containing impact evaluations conducted in low- and middle-income countries. It summarises study findings and methods for researchers, programme managers or policy makers who need information about what works in designing and implementing impact evaluations. The database includes studies that quantify impact using a counterfactual (implicitly or explicitly), with a focus on final welfare outcomes, using qualitative, quantitative or mixed methods.

[21] "Workshop: Impact Evaluation of Trade-Related Policy Interventions: Paving the Way" (*http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/TRADE/0,,contentMDK:22782445~menuPK:2644066~pagePK:64020865~piPK:51164185~theSitePK:239071,00.html*)

[22] Road upgrades in some areas may take longer to achieve. For instance, they may not constitute the most important bottleneck in the region or may otherwise be a low priority.

[23] "International Institute for Impact Evaluation, Synthetic Reviews" (*www.3ieimpact.org/syntheticreviews*).

[24] Examples of synthetic evaluations can be found at *www.3ieimpact.org/systematicreviews*/

References

Agence Française de Développement (2009) *Évaluation des activités de développement – 2009/4. Programme pour le renforcement des capacités commerciales, www.minefe.gouv.fr/directions_services/dgtpe/publi/eval_prcc0906.pdf*.

Banerjee, A.V. and R. He (2008), "Making Aid Work" in W.R Easterly and N. Birdsall (eds.), *Reinventing Foreign Aid*, MIT Press, Cambridge, Massachusetts.

Birdsall, N. (2004), *Seven Deadly Sins: Reflections on Donor Failings*, Center for Global Development Working Paper Number 50, www.cgdev.org content/publications/detail/2737.

Brenten, P. and E. von Uexküll (2009), "Product specific technical assistance for exports – has it been effective?", *Journal of International Trade & Economic Development*, Vol. 18, No. 2.

Cali, M. and D.W. te Velde (2009). "Does aid for trade really improve trade performance?" Overseas Development Institute.

Clemens, M., S. Radelet and R. Bhavnani (2004), Counting Chickens When They Hatch: The Short Term Effect of Aid on Growth, Working Paper 44, Center for Global Development, Washington, D.C., www.cgdev.org/content/publications/detail/2744.

Djankov, S., C. Freund and C.S. Pham (2008), "Trading on Time", *Review of Economics and Statistics*, Vol. 92, No. 1.

Dom, C. (2007), "1. What are the effects of General Budget Support?" *The Joint Evaluation of General Budget Support 1994-2004: Thematic Briefing Papers*, OECD, Paris, *www.oecd.org/dataoecd/14/35/38339258.pdf*.

Duflo, E. and M. Kremer (2008), "Use of Randomization in the Evaluation of Development Effectiveness" in W.R. Easterly and N. Birdsall (eds.), *Reinventing Foreign Aid*, MIT Press, Cambridge, Massachusetts.

Elliot, K. (2007), "Ensuring That Aid for Trade Promotes Development: Indicators for Measuring the Impact," paper prepared for the Workshop on Monitoring and Evaluation, WTO High-Level Conference on Mobilising Aid for Trade: A Global Review, 19 November 2007, Geneva, *www.wto.org/english/tratop_e/devel_e/a4t_e/elliott_e.doc*.

Fitzsimons, E. and M. Vera-Hernandez (2009), *A Practitioner's Guide to Evaluating the Impacts of Labor Market Programs*, World Bank Employment Policy Primer No. 12.

Helble, M., C. Mann and J.S. Wilson (2009), *Aid for Trade Facilitation*, Policy Research Working Paper 5064, The World Bank Development Research Group Trade and Integration Team, *www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2009/09/28/000158349_20090928110357/Rendered/PDF/WPS5064.pdf*.

Lucas, S. (2011) *Millennium Challenge Corporation Principles into Practice: Focus on Results, www.mcc.gov/documents/reports/paper-2011001052001-principles-results.pdf*

OBPR (Office of Best Practice Regulation) (2010), *Best Practice Regulation Handbook*, Australian Government, June 2010, *http://www.finance.gov.au/obpr/proposal/handbook/docs/Best-Practice-Regulation-Handbook.pdf.*

OECD (1991), *Principles for Evaluation of Development Assistance*, Development Assistance Committee, OECD, Paris, www.oecd.org/dataoecd/31/12/2755284.pdf.

OECD (2002), Glossary of Key Terms in Evaluation and Results Based Management, DAC Working Party on Aid Evaluation, OECD, Paris.

OECD (2006), *Trade-Related Assistance: What Do Recent Evaluations Tell Us?*, OECD Development Centre, Paris, *www.oecd.org/dataoecd/31/45/37624158.pdf*.

OECD (2007), "Evaluating Aid for Trade: Proposals for Further Work", OECD, Paris,

OECD (2009), *Managing for Results in Aid for Trade*, OECD, Paris.

OECD/WTO (World Trade Organization) (2009), *Aid for Trade at a Glance 2009 – Maintaining Momentum*, OECD and World Trade Organization, Paris and Geneva.

OECD/WTO (World Trade Organization) (2011), *Aid for Trade at a Glance 2011 – Showing Results*, OECD and World Trade Organization, Paris and Geneva.

Prichett, L. (2009), "Political economy of aid evaluation: How to build a sustainable *and* effective movement", presentation at the conference "What Would the Poor Say? Debates in Aid Evaluation", 6 February 2009, Aid Watch Initiative, New York University,. *http://aidwatchers.com/2009/02/nyu%E2%80%99s-aid-watch-initiative-held-conference-on-%E2%80%9Cwhat-would-the-poor-say-debates-in-aid-evaluation%E2%80%9D*.

Ravallion, M. (2008), *Evaluation in the Practice of Development*, World Bank Policy Research Working Paper 4547, World Bank, Washington, D.C. *www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2008/03/05/000158349_20080305144823/Rendered/PDF/wps4547.pdf*.

Wiebe, F.S. (2008), *Aid Effectiveness: Putting Results at the Forefront: MCC's "New Institutional Approach"*, MCC (Millennium Challenge Corporation), *www.mcc.gov/documents/reports/mcc-112008-paper-results.pdf*.

World Bank (2009), *Unlocking Global Opportunities: The Aid for Trade Program of the World Bank Group*, World Bank, Washington, D.C., *http://siteresources.worldbank.org/INTPREMNET/Resources/UnlockingGlobalOpp.pdf*

World Bank Independent Evaluation Group (2006), *Assessing World Bank Support for Trade, 1987-2004: An IEG Evaluation*, World Bank, Washington, D.C. *http://siteresources.worldbank.org/EXTASSWBSUPTRADE1987/Resources/trade_evaluation.pdf*

Chapter 2
What Can Be Learned from Trade-Related Evaluations?

2.1 Introduction

The OECD meta-evaluation described in Chapter 1 provided an overview of (and perspective on) the ways DAC donors and international agencies have implemented programmes and projects, and how they have conducted their evaluations – in terms of both their methods and the topics covered.[25] The meta-evaluation focused on 162 evaluations, conducted between 1999 and 2010, of programmes and projects in two countries (Vietnam and Ghana) and in the transport and storage sectors in a number of countries. In particular, it looked at whether trade had been an objective of the programmes and projects being evaluated, and at whether trade and development outcomes were evaluated. More broadly, it examined whether the evaluations provided the type of information the trade and development community needs for policy formulations.

The meta-evaluation was based on quantitative analysis (described in this chapter) and qualitative analysis (Chapter 3). To better understand the implicit interests of evaluators, the number of times key words or expressions were used in the 162 evaluations was counted. The meta-evaluation also took a close look at a narrower set of 42 evaluations with a more direct bearing on trade, in order to analyse the content of key words and phrases. That analysis helped to determine how certain core issues were considered.

This dual approach leads to the following conclusions:

- These evaluations of trade-related programmes and projects did not say much about trade. "Trade" and "exports" were not among the most frequently used words. "Imports" were mentioned even less often. References to the World Trade Organization (WTO) and regional or preferential trade agreements were largely absent. Quantitative and qualitative analyses yield similar conclusions. Not only was the trade impact of programmes and projects clearly not the main focus of evaluations, but in a number of cases it was not even addressed.

- The evaluations usually did not clarify the policy linkages which matter most to policy makers. For instance, potential positive or negative impacts of trade policy instruments (*e.g.* tariffs, quotas and subsidies) on funded projects were not assessed or even examined. Nor was there an evaluation of linkages with "behind the border" measures such as regulatory reforms or private sector policies. While some related terms were used in a number of evaluations, qualitative analysis suggests that these terms were not well defined by evaluators, and that their use was not substantiated by reference to hard evidence.

- In contrast to their frequent silence on trade-related issues, the evaluations referred extensively to broad, development-related concepts such as poverty reduction or gender without clearly defining these concepts. A tendency to favour generic concepts over precise terms often meant that evaluations were vague and ill-focused. For instance, gender was either mentioned in passing, as a cross-cutting issue in programmes and projects, or measured in terms of short-term objectives (*e.g.* the number of women participating in a seminar) that had little tangible relevance to a project's impact on women's economic or societal situation.

- Evaluations often lacked an adequate or realistic timeframe for measuring results, rarely distinguishing between what was achievable in the short or longer terms. Following DAC guidelines, most evaluations assessed whether project implementation deadlines were met and budgets were respected, as well as whether the overall programmes and projects were relevant, efficient and sustainable. However, medium- to long-term impacts were not properly measured for particularly important factors such as a programme or project's return on investment. There could be two reasons for this omission. In most cases, little *ex ante* economic analysis of a programme or project had been undertaken, resulting in both a lack of quantifiable definitions of objectives and a lack of baseline data with which to measure impacts. Consequently, in most evaluations there was little economic analysis and sophisticated tools such as economic and econometric modelling were rarely used.

- The evaluators' conclusions gave little insight into whether aid for trade works and why. They were rarely able to identify causal links between programmes or projects and performance. One key problem appears to have been a lack of sustainable financing for many of the programmes and projects surveyed, with the result that they were often terminated prematurely. Another problem appears to have been the difficulty of addressing both economic growth and poverty reduction objectives, as these objectives sometimes conflict. Finally, there was occasionally a problem of poor programme or project implementation, making it difficult (if not impossible) to assess impacts on trade and poverty reduction.

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

It is also possible that the problems described above could pertain only to the relatively small sample of evaluations considered by the meta-evaluation. But none of the donors whose evaluations were analysed appeared to perform systematically better than others, suggesting that an even broader sample of evaluations would not produce substantially different results.

If it is assumed that the results presented here are more or less representative of aid-for-trade evaluations generally, two broad recommendations can be made:

- Every programme and project should include an *ex ante* assessment of the situation being addressed, in order to define objectives in a quantifiable way and provide the information (ideally including data) necessary to measure whether objectives were met *ex post*;

- The sequence of questions in Section 2.3 below could provide a framework for assessing the impact of programmes and projects in a more systematic and thorough way;

    - The quantitative analysis in this chapter and the qualitative one in Chapter 3 point to similar conclusions. One of these is that the word (and expression) counting exercise could usefully be applied in future meta-evaluations.

2.2 Purpose, scope and method of the meta-evaluation

The meta-evaluation had two purposes. The first was to assess whether the selected evaluations provided the information that policy makers need from evaluators. In particular, was trade an objective of the original programme or project, and were trade and development outcomes or impacts taken into account in the evaluations? To the extent that the latter was the case, the meta-evaluation reported on the type of trade, poverty reduction and development outcomes described in the evaluations. It should be emphasised that the aim here was to examine whether these evaluations provided "what we want to know".

The second purpose was to propose a set of guidelines (*e.g.* indicators, quantitative and qualitative techniques) for undertaking future evaluations: in short, to provide a tentative answer to the question "What would we ideally like done?". Two types of guidelines could be proposed that are complements rather than substitutes. The first would be general ones, requiring minimal time, skills and financial resources, which would improve the routine evaluation of any programme or project. The second would be more specific, involving detailed data collection and specific technical skills, and would help undertake more thorough and intensive evaluations of the impacts of programmes and projects. Since the second approach would be much more expensive and time consuming, it would probably be applicable only to a carefully chosen subset of all programmes and projects. For this reason, there is a focus on the first approach.

The meta-evaluation looked at a set of evaluations conducted between 1999 and 2010 of programmes and projects in Vietnam and Ghana and in the transport and storage sectors in a number of countries. There were several reasons for selecting Vietnam: it has served as a model of trade opening for a number of countries; it has enthusiastically embraced the aid-for-trade agenda; it has successfully achieved export-led growth; and it became a WTO member fairly recently, in 2007. 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. The transport and storage sectors were selected because they have received significant amounts of aid for trade and have therefore been subject to a large number of relevant programme and project evaluations (see Chapter 1, Section 1.6).

The set of 162 evaluations includes only those notified to the DEReC database. Thus, not all evaluations undertaken in these countries and in these sectors during this period were used. As a crude indication of the limited or partial scope of the meta-evaluation, the total number of programmes and projects carried out by DAC members between 1999 and 2008 was 61 677.

In the current international aid context, programmes and projects whose purpose is to promote economic development by building trade capacity are often referred to as "aid for trade". In the interests of simplicity, this term is used here to describe all programmes and projects designed to facilitate developing countries' integration into the global economy through expanded trade. Nevertheless, it should be noted that the term became

widely accepted only in the mid-2000s. In view of the time lag between the implementation of a programme or project and its evaluation, it is not surprising that operations and evaluations conducted before the mid-2000s do not refer to "aid for trade" *per se*.

A key challenge in conducting the meta-evaluation was determining the extent to which programmes and projects could be analysed through an aid-for-trade lens. For instance, some of the evaluations in the DEReC database related to programmes and projects in Vietnam and Ghana which were likely to impact trade only in the (very) long run (*e.g.* those related to education or health). Even in the transport and storage sectors, some programmes and projects classified as aid for trade had, at best, remote links to international trade. Therefore, a subset of 42 evaluations which had a more direct and immediate bearing on trade was identified. The overall and narrow sets of evaluations are described further in Chapter 3 and Annex A.

The meta-evaluation was also intended to suggest the implicit interests of evaluators, as well as the gap between these interests and the information needs of aid-for-trade policy makers. This requires both quantitative and qualitative analysis.

The method used for quantitative analysis is comparatively simple. It considers the occurrence of words or expressions that could reasonably be considered crucial in evaluating aid-for-trade related programmes and projects, and that might therefore be expected appear in an evaluation. The 48 key words and expressions selected for the analysis are listed in Table 2.1. This method has the advantage of relying on a clearly defined metric: the frequency with which key words are mentioned per 100 pages (the metric should not be seen as "exhaustive" and "objective", if only because it is limited to these 48 words and expressions). The selected words and expressions do not necessarily serve as an entirely accurate, or even adequate, measure of evaluators' interests or the needs of policy makers. Yet this word set is probably large enough to provide meaningful results and to raise questions about the evaluation process in general.

At the same time, this quantitative approach needs to be complemented by a qualitative one, as the frequency with which words and expressions appear says little about their meaning or the context in which they were used. As described in Chapter 3, such a qualitative reading was undertaken for the narrow set of 42 evaluations, with the idea of looking beyond the issues explicitly treated by the evaluators to learn more about how each issue was being evaluated and assess the significance of the information being provided.

Table 2.1 List of key words and expressions

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

2.3 Quantitative analysis

This section looks at whether the selected evaluations provided information relevant to the aid-for-trade agenda. First, there is a focus on the trade component and on how frequently key words and expressions occur in the evaluations – a crude measure of the revealed interests of the evaluators. There is also an attempt to determine whether patterns emerge. For instance, did a donor use the same concepts in evaluations regardless of the country or sector, which would indicate that there were precise guidelines for carrying out evaluations? Or do the donor's interests appear to have shifted according to the country or sector under consideration, which would indicate the absence of such guidelines?

The trade component

It is reasonable to assume that when policy makers look at evaluations, they are interested in learning about the following main trade aspects of aid-for trade programmes:

- the extent to which evaluations capture the role of trade;
- the extent to which they take into account the trade policy of the country in question;
- the extent to which they refer to trade-related issues specific to aid for trade (*e.g.* technical barriers to trade, sanitary and phytosanitary standards, adjustment policies and trade facilitation);
- the extent to which they take key international trade agreements into account.

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 used the word "trade"? Or have they been more precise by referring to "exports" and/or "imports"? "Exports" and "imports" have very different connotations. Exports would be expected to be a visible trade indicator in aid-funded programmes and projects aimed at better integrating developing countries into the world economy. The aid-for-trade agenda typically focuses on the importance of exports. However, from an economic perspective imports are even 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 generate foreign exchange to pay for much needed imports.

The frequency with which the key words associated with the role of trade appeared in the evaluations is shown in Table 2.2. Three main observations can be made. First, broadly speaking, the trade aspect (defined as the combined occurrence of the words "trade", "exports" and "imports") was relatively prominent (frequencies above 50 per 100 pages were rarely observed in the case of other key words or expressions). Moreover, the frequency with which trade was mentioned was higher in the case of Vietnam than in that of Ghana. Second, there was a greater focus on exports than on imports despite the role of the latter in growth and development. Third, trade terms were used more frequently in the narrow set of evaluations than in the overall set for the two countries and for the transport and storage sectors.

Table 2.2. The role of trade

	Overall set			Narrow set		
	Ghana	Vietnam	Transtor	Ghana	Vietnam	Transtor
How is trade 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 macroeconomic 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

Transtor: Transport and storage sectors combined.

Source: DEReC database. Authors' computations.

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

The second part of Table 2.2 suggests that evaluators paid relatively little attention to clarifying such policy linkages if they were even considered. Specifically, it was not clear whether trade was perceived as a macroeconomic issue or whether macroeconomic imbalances were seen as a trade policy concern. In the narrow set of evaluations there were more references to trade, exports and imports. However, there was little change in how trade and its implicit policy linkages were treated by evaluators.

To what extent do evaluations refer to the trade policies of countries?

Another relevant question is whether evaluators took account of a country's broader trade and economic policy context, especially in the transport and storage sectors, and whether they appreciated its relationship to (and impact on) programmes and projects. For instance, returns from aid-funded programmes and projects could vary greatly, depending on whether a country imposed moderate and stable tariffs or high and volatile ones. Another important variable is whether a country is landlocked. Reducing tariffs would provide only small benefits if transport costs were high and transport infrastructure 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? And 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 quantitative analysis in Table 2.3 concerning the role of trade policy gives rise to the following observations:

- Evaluators almost never used the general term "trade restrictions" despite these restrictions' potential impact on the returns from programmes and projects. This was the case with regard to both the broad and narrow sets of evaluations. It is significant that this observation is valid for the narrow set, given that in the transport and storage sectors a key criterion in selecting the narrow set was the assumption that there would be an explicit international trade dimension in these evaluations (Chapter 3).

- Evaluators addressed the existence of trade instruments only marginally. Tariffs showed a low frequency, and subsidies an even lower one, although most aid-funded programmes and projects in the transport and storage sectors ultimately have a substantial subsidy component. Evaluators tended to limit their analysis to a comparison of subsidy schemes provided by various countries. They stopped 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 (Chapter 3).

- Evaluators gave little attention to 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 by domestic and foreign firms alike.

- More attention was given to these instruments in the case of programmes and projects in Vietnam, which is in the process of undertaking much faster and deeper liberalisation than Ghana. This may reflect the rapid evolution of Vietnam's trade policy during the

period, whereas Ghana's was evolving less rapidly. At the same time, it raises serious questions about the usefulness of evaluations if they do not address existing policies that could have a direct bearing on the success of programmes and projects.

Table 2.3. The role of trade policy

	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

Transtor: Transport and storage sectors combined.

Source: DEReC database. Authors' computations.

Again, donors showed parallel behaviour. In particular, evaluators paid little attention to trade restrictions in the two countries or in the transport and storage sectors. Variances were higher for tariffs and subsidies, with the African Development Bank (AfDB) and the Asian Development Bank (ADB) acknowledging the importance of tariffs in Vietnam and Ghana and only the Agence Française de Développement (AFD) highlighting the relevance of subsidies. The transport and storage sectors are not comparable because tariffs are often used as an equivalent of price.

To what extent do evaluations refer to trade-related issues?

The aid-for-trade agenda covers a much wider set of trade-related issues than the basic trade instruments examined above. They include: (i) technical barriers to trade (TBT) for industrial goods; (ii) sanitary and phytosanitary (SPS) standards for food safety and animal and plant health; (iii) trade facilitation; (iv) trade-related adjustment policies; and (v) trade-related assistance. Table 2.4 shows that in the set of 162 evaluations these issues were by and large ignored. However, in the narrow set there were more frequent references to SPS standards (in Vietnam), trade facilitation (in Vietnam and the transport and storage sectors), adjustment policies (in Ghana) and trade assistance (in Vietnam).

Table 2.4. The wider set of trade issues

	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

Transtor: Transport and storage sectors combined.

Source: DEReC database. Authors' computations

Some of these results are puzzling. The almost total absence of references to TBT, SPS standards and trade-related technical assistance is particularly surprising, given the priority typically given them by both recipients and donors. It should be noted that the relative frequency with which trade facilitation was mentioned in the case of Vietnam is consistent with its being the world's second largest recipient of aid related to trade facilitation between 1999 and 2008 (65 projects or 2.6 per cent of all projects).[26] However, it is surprising that in the narrow set adjustment policies were by a significant margin mentioned more frequently in the case of Ghana than in that of Vietnam. During the decade examined, particularly at the end of the decade, Ghana had a more stable product composition with regard to exports and imports than Vietnam. Heightened sensitivity to adjustment issues in Ghana could reflect the long debate in many African countries about donors' adjustment programmes of the 1980s and 1990s.

Do evaluations take key international trade agreements into account?

The aid-for-trade agenda has developed largely independently of the Doha Development Round and negotiations of regional trade agreements. Nevertheless, trade negotiations have dominated the debate over the last decade. Thus, it is useful to look at whether key trade agreements (in particular, the Doha Round of multilateral negotiations) were mentioned in the evaluations. References to regional or preferential trade agreements would be significant since such agreements have echoed, and often amplified, WTO discussions on aid for trade.

Table 2.5 shows that there was an almost complete lack of any mention of key trade agreements in the evaluations, with the exception of references to the WTO. Even the WTO references were almost entirely limited to programmes or projects in Vietnam, whose ongoing WTO accession negotiations were a key policy issue during the decade covered.[27] The absence of references to regional or preferential trade agreements in Ghana's evaluations may reflect that country's resistance to accepting new commitments during the Doha negotiations. It could also reflect Ghana's relatively limited involvement in regional or bilateral agreements, as well as the modest impact of these agreements on its trade policy. For instance, the Economic Community Of West African States (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 transport and storage sectors, the absence of references to trade agreements exemplifies the basic difficulty of negotiating services liberalisation (particularly in the case of transport services, which have long been subject to special international schemes geared towards limiting international competition) in most trade agreements.

There were few references to trade agreements in either the overall or narrow sets. The one notable exception was the high frequency with which the WTO was mentioned in EU and Swiss Secretariat for Economic Affairs (SECO) evaluations of programmes and projects in Vietnam.

Table 2.5. International trade agreements

	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

Transtor: Transport and storage sectors combined.

Source: DEReC database. Authors' computations.

The development component

When the focus shifts from issues related to trade to those more closely related to development, three questions seem particularly relevant:

- Do evaluators take into account the constraints that an inefficient public sector could impose on development?
- Do they take into account the broad constraints related to inefficiencies in the private sector?
- Do they 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. These constraints can be even more difficult to overcome in developing countries than in developed ones. From a development perspective, it is essential for evaluators to give some sense of their perception of whether public sector generated constraints exist, and whether such constraints have had a substantial impact on the returns from the programmes and projects evaluated. Against that background, this section looks at another spectrum of policy linkages, among aid-funded programmes and projects, 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 programmes and projects that might impact on the success of the ones evaluated? Second, do evaluators focus on regulatory infrastructure issues that might also impact on the success of a programme or project? To operate efficiently, firms need appropriate domestic regulations as well as appropriate roads or bridges. The role of this regulatory capital can be as important as the role of physical or human capital in economic development.

Table 2.6 suggests how these various public sector-related constraints were understood by evaluators. The overall and narrow sets give similar results, with the usual magnification effect in the narrow set. It is clear that evaluators were concerned with physical infrastructure problems. The systematically higher frequency with which physical infrastructure was mentioned in the transport and storage sectors compared with

the two countries suggests that evaluators were even more inclined to take a broad view of the interactions between a programme or project and other physical infrastructure when they were already involved in sectoral infrastructure policy. However, the lower frequency in the case of the two countries could simply reflect evaluators' inability to take a bigger picture view of economies as a reflection of multiple, interactive operators (*i.e.* a "general equilibrium" perspective).

In sharp contrast, evaluators showed less interest in some regulatory issues, whether specific or general. In the table, regulatory policies are divided into three subsets: (i) specific regulations closely related to the evaluated programmes and projects; (ii) the broad regulatory framework of the country or sector; and (iii) the even broader issue of "governance". For Ghana, in particular, references to specific regulations were relatively infrequent while there were almost no references to the broad regulatory framework. 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 had little to say in detail about governance. Finally, regulatory issues (specific and general) were more salient in the case of Vietnam than in that of Ghana. This may give the impression that Vietnam's regulatory policies are somewhat better defined than Ghana's. On the other hand, broad governance issues were given more attention in the case of Ghana than in that of Vietnam.

Table 2.6. Public sector-related constraints

	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

Transtor: Transport and storage sectors combined.

Source: DEReC database. Authors' computations.

How much attention is paid to supply-side constraints?

The literature on development gives considerable importance to supply-side constraints in the economic growth equation. Since public sector-related constraints are only one aspect of supply-side constraints, this section focuses on private sector-related constraints. However, it is acknowledged 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 emphasises the importance of supply-side constraints, evaluators might have been expected to focus extensively and systematically on this issue. They might also have been expected to provide information on expenditure or costs, either the direct costs involved in completing the programme or project or, more interestingly from the perspective of this section, the indirect costs to the whole economy of not (or only partially) completing a programme or project. In this context, evaluators might also have been expected to provide some sense of whether the private sector had grown as a result of a programme or project. Finally, it might be asked whether evaluators showed a systematic interest in technical assistance in the broader sense of the term (*i.e.* not just trade-related technical assistance).

Table 2.7. Supply-side constraints

	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

Transtor: Transport and storage sectors combined.

Source: DEReC database. Authors' computations.

The quantitative analysis in Table 2.7 presents a mixed picture. Evaluators did not address supply-side constraints in the overall or narrow sets of evaluations. There was a greater focus on expenditures (costs), but only in the case of Ghana were these mentioned frequently and then only by four donors, the Danish International Development Agency (DANIDA), the UK Department for International Development (DFID), the EU and the World Bank. Evaluators used the term "expenditures" in two ways: in the context of expenditures exceeding budgets; and to differentiate between a programme or project's recurring expenditures (often assumed to be covered by the recipient) and non-recurring expenditures (often funded by the donor).[28] The term "private sector growth" rarely appeared in either the overall or narrow sets of evaluations. The term "private sector" appeared much more often, reflecting its frequent use in neutral statements (*e.g.* "encouraging the development of the private sector by enhancing the business environment in which it operates") rather than in the context of evaluating specific programmes or projects (Chapter 3).

How much attention is paid to performance?

The basic aim of evaluations is to provide a sense of whether programmes and projects or (general aid policies) have performed according to expectations. Given that development is a broad subject, it is difficult to compile a concise or exhaustive list of words and expressions that reflect evaluators' concerns about development outcomes. However, the following terms – ranging from the more precise to the more general – were identified as key to such a list: "economic growth", "competitiveness", "efficiency", "effectiveness", "sustainability", "poverty reduction" and "gender gap".

The frequency with which key development-related words appeared in evaluations of programmes and projects is shown in Table 2.8. One clear observation emerges: more general words and expressions were used much more frequently than more precise ones. On the surface, this observation is puzzling given that it would presumably be more difficult for evaluators to describe a programme or project using general words and expressions like "poverty reduction" or "gender gap" than more precise ones like "economic growth. The answer, as illustrated by the qualitative analysis (Chapter 3), may lie in evaluators' typically using these terms to describe a programme or project's objectives rather than to assess its performance in meeting those objectives.

Table 2.8. Development performance

	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

Transtor: Transport and storage sectors combined.

Source: DEReC database. Authors' computations.

2.4 Conclusions

The quantitative analysis of 162 evaluations of programmes and projects in Vietnam and Ghana and in the transport and storage sectors in a number of countries demonstrates that there were few clear links between donors and specific issues, that is, there were few systematic "leaders" in terms of evaluations of a given topic for a given country and/or sector. A donor might pay attention to a particular issue in the case of one country or sector, but completely ignore the same issue when considering another country or sector. Systematic evaluations should thoroughly examine all potential issues and topics. If a particular key issue or topic is ignored, the reasons for doing so should be clearly stated. This interpretation of the data argues strongly for a serious effort to establish more detailed guidelines and a stricter system of reviews for evaluations. In turn, this might allow greater freedom and creativity in evaluations.

The key words and expressions that are most precise, from an economic analysis perspective, tended to appear less frequently in evaluations. Moreover, evaluations gave little information on the broader economic context in which programmes and projects were taking place (the "general equilibrium" perspective), making it difficult to clarify key policy linkages. This was particularly striking in the case of the linkages that would have been easiest to assess. For instance, when evaluators ignored trade policy instruments, this suggests they had little understanding of – or gave little weight to – the extent to which programmes or projects' rate of return could be reduced or increased through the use of such instruments. High tariffs on inputs crucial to a programme or project would obviously limit its effectiveness, just as high tariffs imposed on foreign substitutes for a programme or project's output could artificially increase its benefits. Including this kind of analysis in evaluations would be particularly useful for recipients, allowing them to 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 programmes and projects).

Notes

[25] The evaluations analysed by the OECD meta-evaluation had been notified to the OECD DAC Evaluation Resource Centre (DEReC) database.

[26] Source: OECD/DAC CRS.

[27] Vietnam's negotiations for accession to the WTO started in January 1995 and were completed in November 2006. It is undergoing a long and as yet uncompleted process of implementing its commitments.

[28] The term "expenditures" was used to refer to public expenditures. This is particularly true in the case of the transport and storage sectors, where several evaluations stressed the private sector's lack of involvement in managing infrastructure.

Chapter 3
Do Aid-for-Trade-Related Evaluations Tell Us More?

3.1 Introduction

The qualitative analysis described in this chapter is largely consistent with the conclusions reached through quantitative analysis in Chapter 2. The impact of programmes or projects on trade was clearly not the focus of evaluations. In a number of cases the trade impact was never explicitly addressed. While the use of some terms suggests that a number of evaluations did give attention to policy linkages to trade, the qualitative analysis demonstrates that these terms were not well defined. Nor was their use substantiated by reference to hard evidence.

The evaluations focused more on broad, development-related concepts than on precise trade issues. Even 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 objectives that had little meaningful bearing on women's economic or societal situation (*e.g.* the number of women who attended a seminar).

Evaluators rarely incorporated relevant time dimensions in their assessments. Two main reasons for this can be suggested. First, little *ex ante* economic analysis of the programmes and projects had been undertaken, meaning that their objectives remained unquantified and there was no baseline data with which evaluators could measure their impact. Second, and to a great extent consequently, little economic analysis was undertaken in most evaluations.

A quantitative analysis of key words and expressions works best when there is a large set of evaluations. However, this chapter presents a qualitative analysis of the narrow set of 42 evaluations (with the addition of a 2010 evaluation of a project in Vietnam, which has been included because of its special features). The main purpose of this analysis is to test the observations made in Chapter 2. With the goal of applying coherent methods across country-related and sector-related evaluations, a series of questions was applied to the narrow set of evaluations to explore the characteristics of each programme or project and assess the evaluations' outcomes.

The framework was adjusted, depending on the purpose of the programme or project and whether it had a direct aid-for-trade orientation. For Vietnam, Ghana, and the transport and storage sectors the following sections explain the characteristics of the selected evaluations, the methods used, and the issues and topics covered. The overall aid-for-trade impact of the programmes or projects is then assessed.

3.2 Is aid for trade working in Vietnam?

The selected evaluations

Of the 71 evaluations specific to Vietnam (or to a set of countries including Vietnam), the 10 most relevant were selected for the narrow set and classified according to the WTO Aid for Trade Task Force's definition of aid for trade, *i.e.* trade-related technical assistance (TRTA) and infrastructure and private sector development. Programmes or projects focused on the domestic market were not included, even if they fell into the private sector development category (*e.g.* aimed at producing paper for the Vietnamese market).

Trade did not figure nearly as prominently in evaluations of private sector development programmes and projects as in evaluations of TRTA and infrastructure programmes and projects. A qualitative analysis of Vietnam's narrow set of evaluations shows that only comprehensive reviews looked at trade enhancement and poverty reduction in any detail. This was mainly because these reviews were concerned with long-term impact assessments, whereas the focus of programme and project evaluations tended to be more short-term. Such reviews were only carried out by the Asian Development Bank (ADB).

All the selected evaluations gave positive ratings to programmes and projects, even if some also mentioned aspects that were less satisfactory. There were notable differences among the different categories of programmes and projects evaluated. Some assessments were based solely on field missions (consisting of interviews with randomly selected stakeholders), whereas others also made use of objective indicators such as rates of return. The extent to which donors had harmonized their programmes and projects with each other and with governments' objectives and strategies was also usually noted – and positively rated – in evaluations.

Methods and focus

The methods used by evaluators were highly dependent on the type of programme or project being assessed. For instance, quantitative tools were barely used in assessing TRTA programmes or projects, but were used extensively to evaluate infrastructure programmes and projects. Two general comments apply to all the selected evaluations. First, a programme or project's impact on entrepreneurs, exporters and importers was hardly ever evaluated, with the exception of a few case studies in a Swiss Secretariat for Economic Affairs-United Nations Industrial Development Organization (SECO-UNIDO) evaluation. According to many evaluators, this reflected a failure to collect relevant data at the time of the programme or project's implementation, as well as the insufficient time between the end of programmes or projects and the start of evaluations, especially TRTA evaluations. Second, few evaluations used econometric models to assess the impact of infrastructure programmes or projects on poverty reduction. The one exception was not really an evaluation, but rather a prospective analysis. In none of the other evaluations were econometric tools used to establish causality. Evaluators frequently mentioned the need for more robust analysis in their evaluations. Only the ADB reviews addressed trade or socio-economic impacts, and here the impact assessment was usually subjective.

Trade-related technical assistance

Four TRTA programmes or projects were implemented to strengthen institutional reform in Vietnam, as part of the country's effort to open its economy and accede to the WTO. Each focused on setting up workshops or developing certification and testing facilities, 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 the Swedish International Development Cooperation Agency (SIDA). The fourth evaluation assessed an EU-funded programme.

All of these evaluations were carried out by independent consultants. The consultants followed donors' guidelines (which were similar across donors) and analysed the programmes and projects' impact, relevance, efficiency and sustainability, as well as their effect on a series of broader cross-cutting issues such as gender and environment. All of the evaluations were qualitative, based mainly on field missions. Evaluators frequently complained about their inability to assess the outcomes of programmes and projects more thoroughly because of the lack of baseline data against which progress could have been measured. This clearly suggests that quantifiable programme or project objectives should be clarified from the start, when basic information and data that would allow the assessment of progress, outcomes and impacts should also be collected.

Without this basic information and data, evaluators assessed outcomes according to two more subjective criteria: the effective implementation of the project (*e.g.* the number of participants at workshops, a poor indicator of the capacity to apply acquired knowledge) and stakeholder interviews. Outcomes were never objectively assessed in terms of effective knowledge transfer. Furthermore, the causal link between the various TRTA programmes and projects and trade development was never clearly established other than by relying on the evaluators' opinions. In fact, evaluators often used the expression "in our view".

Infrastructure programmes and projects

All three evaluations assessed had been carried out by the ADB. Two were related to the same regional programme. One was a working paper by independent authors from the ADB Institute, while the other two were undertaken by the ADB's Programmes and Evaluation Department. Generally, these evaluations employed more robust impact assessment methods. In assessing the impact of infrastructure projects, ADB evaluators collected primary data (*e.g.* traffic or border-crossing point counts) and undertook qualitative surveys to assess the impact of trade facilitation projects on the movement of goods and people across borders.

The evaluation of the rehabilitation of the Port of Ho Chi Minh provides a good illustration. Evaluators were able to forecast cargo handling performance, and the evolution of various other flows, based on whether or not the project was implemented. Use of this approach was possible because key indicators had already been defined and were available as part of the initial project outline. However, the evaluation team was not able to determine the extent to which these economic benefits would be passed on to the Vietnamese economy as a whole.

As highlighted in the quantitative analysis, evaluators tended to focus more on exports than imports. In this respect, the ADB infrastructure reviews and evaluations are

exceptions. The evaluation of the Port of Saigon's rehabilitation, for instance, highlighted the importance of imports for Vietnamese businesses. It pointed out that a decrease in the import prices of key inputs such as fertilizer, iron and steel was crucial to the development of the country's economy. Evaluations of programmes or projects in the transport and storage sectors also emphasised the importance of such a decrease.

The ADB review of the Greater Mekong Sub-region programme (Stone, Strutt and Hertel, 2010) deserves special mention. It was one of the most thorough evaluations covered by the meta-evaluation. Not only was the timeframe long enough for evaluators to quantify impacts, but various impact assessment studies directly relevant to the programme were included. Although the impact assessment studies had more than one focus (*i.e.* socio-economic, trade, investment, poverty reduction), they provided valuable information, based on robust methods, which combined offered a comprehensive assessment of the programme's overall impact.

Three factors could explain why evaluators used quantitative tools more extensively in evaluations of infrastructure programmes and projects than in those of TRTA:

- The timeframe for infrastructure programmes and projects is much longer than for TRTA (the two ADB infrastructure projects covered more than ten years), meaning that evaluators have scope to employ more sophisticated assessment tools.

- Donors are keen to ensure a positive rate of return on their investments in large-scale infrastructure projects. They typically insist on the importance of gathering *ex ante* information and using rigorous methods when evaluating the impacts of these projects.

- The outcomes of such programmes and projects (*e.g.* whether the number of cargo ships using new harbour facilities increased) are easier to monitor and evaluate quantitatively than those of TRTA (*e.g.* whether entrepreneurs and/or officials used knowledge gained in workshops).

Private sector development

Three evaluations relating to programmes or projects for private sector development, undertaken by the ADB, German Technical Co-operation (GTZ) and Sida, were selected from the available evaluations. The qualitative analysis of these private sector development evaluations confirms earlier conclusions based on the results of counting words and expressions (Chapter 2). In none of the evaluations, with the possible exception of that by the GTZ, was trade explicitly considered nor was the impact on poverty ever assessed, even in the longer-term reviews.

The ADB evaluation consists of a wide review of all its private sector development programmes and projects carried out between 1985 and 2006. Vietnam is one of several case studies. The focus of the evaluation is on performance and strategy, not on impacts. The assessment of a programme or project relies mostly on an analysis of investment performance. Only a few details are provided concerning one ADB project that created jobs and had an overall positive socio-economic impact. Although the GTZ and SIDA evaluations relied mostly on qualitative information, they also contained robust quantitative analysis and there was a comprehensive attempt to assess longer-term

impacts. For instance, the SIDA evaluators mentioned not only the direct beneficiaries of the private sector development programme but also the "ultimate" intended beneficiaries, meaning small- and medium-sized enterprises (SMEs), although they were unable to assess the impact on these beneficiaries. Evaluators also noted the importance of allowing sufficient lead time for entrepreneurs to absorb and benefit from programmes and projects before attempting to assess their impacts, and the impossibility of drawing useful conclusions when timeframes are too short. Consequently, the evaluators felt unable to infer that growth of SMEs was a consequence of any specific programme or project outcome. The same problem affected the GTZ evaluator's capacity to assess the impact of programmes or projects aimed at private sector development.

Outcomes of the evaluations

Reaching clear conclusions about what works and why in aid for trade remains difficult. Existing evaluations tend to tell us much more about how a programme or project was implemented and its economic sustainability (*e.g.* with information on rates of return) than about how it impacted on trade or poverty reduction.

It is important to highlight once again the uniqueness of Vietnam in the OECD meta-evaluation, in terms of the clear linkages between its WTO accession and the aid-for-trade agenda and the Vietnamese government's firm commitment to trade and business development. In short, Vietnam served an ideal "laboratory" for the aid-for-trade agenda.

Despite the positive synergies in Vietnam, however, evaluators were unable to establish an impact assessment concerning the impact of aid for trade on trade performance, much less its impact on poverty reduction. Although the relevance of many programmes or projects in this context was beyond doubt, it remained difficult to infer any direct causality between aid-for-trade programmes and projects and Vietnam's positive economic and trade dynamic.

3.3 Is aid for trade working in Ghana?

The selected evaluations

Of the 41 evaluations available for Ghana in the DEReC database, 7 were not used either because they were not *ex post* evaluations (but rather *ex ante* examinations of programmes or projects conducted to determine whether they should be financed) or because they did not directly concern Ghana. Of the remaining 34 evaluations, 13 (spread equally across programmes, projects and aid policies) were selected for the qualitative analysis. Given the small number of evaluations available, and their tenuous links to trade, the selection procedure was based on a simple criterion: if the purpose of a programme or project was to facilitate the integration of developing countries into the global economy through the expansion of trade (even if this objective was marginal), the evaluation was selected. Among the 13 evaluations selected, only a few can be said to evaluate an aid-for-trade programme or project strictly speaking.

The broad definition of aid for trade used in selecting the narrow set of evaluations for Ghana is very different from the definition used in the case of Vietnam and the transport and storage sectors. Therefore, the data can be compared only with caution. To some extent, the lower frequency of trade-related terms in the Ghana set is to be expected since the evaluations were not directly concerned with aid for trade. Half of the 13

programme or project evaluations assessed were strongly biased towards aid policies. To the extent that the evaluations in the DEReC database are representative of evaluations of programmes or projects in Ghana, it seems clear that donors did not focus on aid-for-trade programmes or projects in that country. This might be due to the poverty reduction priorities of the government and of some donors. However, it is difficult to reach firm conclusions. Since there is no complete list of the programmes and projects donors have carried out and evaluated, the representativeness of the DEReC database cannot be assessed.

Methods and issues covered

Most evaluations closely followed the DAC guidelines. However, the DAC criteria were used to look mainly at short-term outcomes of the programmes and projects. Techniques that would have allowed robust performance assessments were largely absent. Most evaluations assessed the relevance of programmes and projects with regard to whether they fit the country context, addressed the government's objectives and strategies (*i.e.* policy linkages) and met budget, disbursement and implementation targets. Impact assessments were often hypothetical and lacked any objective measurement. None of the evaluations used quantitative tools, and the evaluation processes never involved data collection, construction of indicators, or quantitative techniques.

Lack of data was frequently mentioned by evaluators 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. Based on evaluators' descriptions of their methods, they relied exclusively on one or both of the following tools:

- analysis of documents obtained from officials and sector stakeholders;
- analysis of the results of field interviews with relevant stakeholders and/or focus groups.

These tools were also used informally in evaluations where no particular methods were identified.

Such practices are problematic in two respects. First, it is difficult, if not impossible, to assess outcomes exclusively through use of qualitative methods. Second, use of qualitative methods can be challenging. Not only are key documents often unavailable, but it is rarely explained how the available documents were systematically analysed or how results were derived. In some evaluations economic theory was used to predict the possible impact of programmes and projects, 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 programme or project.

The elusive quest for trade

Most evaluations investigated the impact of programmes and projects on poverty reduction, the environment, social and gender issues, and (as noted above with regard to Vietnam) private sector participation. In contrast, the impact on trade was not always explicitly considered. Although all 13 of the evaluated programmes and projects dealt with trade-related issues (*e.g.* supply-side constraints or transport and energy market issues), none was explicitly designed to achieve trade objectives. Moreover, when the impact on trade was mentioned in evaluations, it was most often referred to in theoretical or hypothetical terms and never systematically measured or verified.

The word "trade" occurred on average 13 times per 100 pages (Table 2.2), half the number of times as "poverty reduction" (Table 2.8). This gives a good idea of the evaluators' limited focus on trade. In addition, references to trade did not occur in trade impact analysis, but rather in the country context provided by the evaluators. One of the "aid policy" evaluations included a paragraph explicitly evaluating the impact of several programmes or projects (*i.e.* those related to transport) on international trade. It also contained a few paragraphs examining the impact of programmes or projects on private sector development and on the diversification of Ghana's productive/export base. These paragraphs were relatively short and, like the other selected evaluations, this one did not contain quantitative impact measurements.

Outcomes of the evaluations

The selected evaluations did not offer a definitive answer to the question of whether (or why) aid for trade worked in Ghana. This is because evaluators were unable to identify causal links between programmes or projects and performance. The results reported in evaluations with regard to funds disbursed and programmes or projects implemented appear to have been mixed. One recurring problem appears to be the sustainability of programmes or projects. Several evaluations reported positive results in the short term, but less positive ones over the longer term because of financial uncertainty.

Evaluators mentioned the difficulty of tackling growth and poverty reduction objectives simultaneously. For instance, in one evaluation support to the agricultural sector was said to have a positive pro-poor impact through successful targeting of poor recipients although it failed to promote growth or export diversification, raising serious questions about its sustainability. The same evaluation found that support to the transport sector had a positive impact on growth but not on poverty reduction, as it was targeted mainly at the richest regions.

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

3.4 Is aid to the transport and storage sectors improving trade performance?

The transport and storage sectors are by nature trade oriented. All programmes or projects devoted to rehabilitating, maintaining or constructing transport infrastructure (*i.e.* roads, railways or ports) should result, at least indirectly, in enhanced trade and economic growth as well as poverty reduction. Transport costs account for a large part of the price of imported and exported goods. Especially in African or landlocked countries, transport costs seriously weaken the terms of trade.[29] The World Bank (2007) emphasised that the effects of transport investment on poverty reduction are not well understood, although "*growing evidence links transport investment to the improved well-being of the poor*" and, even if they are largely indirect, "*most direct poverty-targeted interventions such as schools, clinics, or nutrition programmes depend on transport in one way or another*".[30]

The selected evaluations

Of 66 relevant studies and evaluations in the DEReC database, 20 were selected for the narrow set of evaluations because of their international trade dimension. The proportions of donors in the overall and narrow sets are comparable, with the ADB responsible for a majority of the evaluations. Programme or project implementation dates were largely during the period 1995-2000. Several evaluations refer to specific projects, while others refer to a programme. Most transport infrastructure programmes and projects pertain to the road sector. The diverse nature and quality of this sector is well represented in these evaluations, as programmes and projects targeted national and local roads, expressways, transit roads, and even entire networks. In the maritime sector, programmes and projects focused on port maintenance or construction. Those focusing on bridges, railways and airports were marginal in both the overall and narrow sets of evaluations. The narrow set covers a balanced range of recipient countries, even if Asian countries are heavily represented.

Methods

An analysis of the selected evaluations leads to a number of observations. First, transport and storage programmes and projects did not target specific populations. Although many of them impacted on rural populations, often in agricultural areas, this was not among their objectives. The same is true with regard to the impact on women. Second, a programme or project's success was always assessed at the programme level. In country-related analysis, different dimensions were assessed, such as outcomes, relevance, effectiveness, efficiency, institutional impact, socio-economic impact, environmental impact, and sustainability or cross-cutting impacts.

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 and storage sector in question (best illustrated by the ADB's evaluations). Despite describing initial conditions, evaluators often expressed dissatisfaction with the lack of baseline data to use in validating the impact of programmes or projects.

Concerning methods, there were some disparities among evaluations. Delays and budgets were evaluated in all of them. While all attempted to provide some sort of rate of return (economic internal rate of return, EIRR, or economic rate of return, ERR) assessment, only a few included a cost-benefit analysis. Several evaluations tried to undertake a cost-benefit analysis by assessing alternative scenarios and providing sensitivity analysis.

Almost all the selected evaluations gave a positive rating to outcomes, in that programmes or projects were usually considered to have been 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 steadily improved since the early 1990s and have been higher than those in other sectors.

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 a programme or project's objectives, trade was always mentioned by evaluators as one of the expected results. However, when it came to measuring trade impacts, figures were hard to come by and evaluators acknowledged the limitations of their analysis.

The universal indicators used in transport and storage sector evaluations are vehicle operating costs (VOC), traffic flows and time saving (Nordic Consulting Group, 2008). In the case of port programmes or projects, the basic indicator is port throughput, a series of efficiency indicators related to the port's productivity. These indicators could provide a useful basis for assessing the impact of transport-related improvements to trade, but no extrapolation was undertaken in any of the 20 evaluations. When a systematic measurement of transport costs proved impossible, some evaluations attempted to approximate the evolution of transport costs following implementation of the programme or project. Generally, when trade was assessed there was a greater focus on exports than on imports. However, there were several notable exceptions, with some attempts to assess the programme or project's likely impact on the costs of imports. This information was mostly obtained through interviews.

With regard to private sector impact, evaluations largely acknowledged the key link between transport infrastructure improvements and private sector development. Several evaluations tried to measure impacts on the creation of businesses (especially SMEs), industrial expansion or even growth of tourism. Increased employment or, more interestingly, the net balance of jobs created and destroyed was also measured in some evaluations.

Outcomes of the evaluations

The evaluations found that the trade impact of programmes or projects was positive, based mostly on growth of traffic volumes. However, these conclusions were reached without taking into account counterfactuals or the effects of other external factors. Some evaluations recognised the synergies to be realized from donors working together. Assessments of these synergies and other potential externalities are difficult to quantify and, at a minimum, require a certain level of agreement on methods among donors. This is especially true for a complex sector like transport, which requires coherence across modalities (*e.g.* road, railways) and across territory to deliver expected benefits. In particular, close attention needs to be paid to trans-modal coherence in any evaluation. For instance, it makes little sense for a programme or project to target just one aspect of transport infrastructure, such as road maintenance, without reference to the broader transport network and the regulatory environment of which it is a part. This concern is more adequately addressed in evaluations that attempt to assess a number of transport sector programmes or projects together and comprehensively.

3.5 The evaluation process

It is useful to look at the evaluation process as revealed in the evaluations themselves. This can be done by:

- looking at the evaluation procedures used;

- looking at the economic instruments used for evaluating programmes and projects in other domains;

- looking at the econometric instruments used for evaluations in other domains.

A frequent complaint by evaluators was that they lacked an *ex ante* assessment of the programme or project. Objectives tended to be defined before the work required to make the programme or project assessable had been carried out. In this respect, evaluations in the transport and storage sectors were generally of better quality than evaluations of programmes and projects in Vietnam or Ghana because certain technical or engineering work had been done before a programme or project was launched and this preliminary work required a clear definition of objectives and means. It is revealing that the word "objective" appeared in evaluations much more often than the word "outcome".

Against this background, it is useful to look at how evaluators perceived their own procedures. Table 3.1 shows the frequencies with which four procedural terms appeared in evaluations. "Indicator(s)" and "performance(s)" are concerned with outcomes, while "monitoring" and "review" relate to procedures that are critical to ensure the quality of aid-for-trade related programmes and projects. There was a notable variation in the frequencies of the four key words, with generally lower frequencies in the case of Vietnam than in that of Ghana.

Table 3.1. Key evaluation procedures

	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

Transtor: Transport and storage sectors combined.

Source: DEReC database. Authors' computations.

Shifting from procedural to more substantive issues, Table 3.2 shows the frequency with which several terms that are part of routine evaluation work in other domains were used by evaluators. "Impact assessment(s)" 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 EU) to the *ex post* impact of regulatory changes that have already been introduced (the United States). "Cost efficiency" and "cost-benefit" refer to well known techniques (with different strengths and weaknesses). "Short term", "long term" and "discount rate" are also included; taking adequate time horizons into account can play a key role in ensuring 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 must limited resources carefully over time.

The table shows that these terms were little used by evaluators, with the exception of references to the time horizon. "Long term" was used more often than "short term" However, on its own "long-term" has little meaning without an awareness of the discount rate and its role in long-term cost-benefit analysis. Here the overall and narrow sets of evaluations do not differ greatly.

Table 3.2. Key econometric techniques (1)

	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

Transtor: Transport and storage sectors combined.

Source: DEReC database. Authors' computations.

Sophisticated techniques are increasingly used in other aid domains where public choice is involved (*e.g.* education and health).[31] To assess whether these techniques were used in evaluations, Table 3.3 shows the frequencies with which four key terms ("counterfactual(s)", "control variable", "differences in differences" and "randomisation") appeared. Almost all of these terms were totally absent in both the overall and narrow sets.

Table 3.3. Key econometric techniques (2)

	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

Transtor: Transport and storage sectors combined.

Source: DEReC database. Authors' computations.

3.6 Conclusions

The impact of programmes and projects on trade was clearly not the focus of evaluators' work. In a number of cases this impact was not addressed at all. In others it was addressed only in very general terms, while the impact on exporters and/or importers was hardly ever evaluated. Simple quantitative tools, such as rates of return, were sometimes used, especially when infrastructure projects were analysed. However, except in a few instances (*e.g.* evaluations of programmes and projects in the transport and storage sectors) the use of more sophisticated econometric tools was rare, mainly because of a generalised lack of information and data collection, both *ex ante* (during programme or project implementation) and *ex post*.

Outcomes differed from one case to the next. While all the selected evaluations of programmes and projects in Vietnam generally gave positive ratings, the selected evaluations of those in Ghana gave mixed ones. This difference might stem from the very different nature of the programmes and projects under consideration. In evaluations of programmes and projects in both countries, donor harmonisation and coherence with governments' objectives and strategies were usually mentioned and positively rated in the evaluations. In contrast, the sustainability and impact of the programmes and projects evaluated were harder to document and their outcomes seemed less satisfactory.

Notes

[29] According to the World Bank (2007), in Africa 11.5% of the total value of imports was related to transport costs and 20% to export costs (up to 55% in landlocked countries such as Malawi). There is extensive literature on this topic, of which relevant examples are found in Djankov, Freund and Pham (2006), Hummels (2007) and Raballand and Macchi (2008).

[30] The World Bank (2007) provides a series of examples through research projects and surveys. Other papers on this topic are cited in Stone, Strutt and Hertel (2010).

[31] For an excellent survey, see World Bank (2007).

References

Djankov, S., C. Freund and C. Pham (2006), *Trading on Time*, Policy Research Working Paper Series 3909, The World Bank, Washington, D.C.

Hummels, D. (2007), "Transportation costs and international trade in the second era of globalization", *Journal of Economic Perspectives*, Vol. 21, No. 3.

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

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

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*.

Chapter 4
Getting Results in Aid for Trade: Setting Targets and Using Performance Indicators

4.1 Introduction

The 2005 Paris Declaration on Aid Effectiveness put the results agenda firmly in the centre of global efforts to improve aid effectiveness. The growing focus on performance is centred on strong notions of goals, causality and continuous improvement. Evaluation is an integral part of this process. It provides important sources of performance information that can contribute to management learning and improved decision-making processes.

Against this backdrop, bilateral and multilateral donors are increasingly putting in place essential building blocks for results-based management to ensure that their activities achieve the desired objectives and targets. They manage *for* results through articulating a results chain from project inputs, to activities, outputs, outcomes and long-term impacts. The results chain provides a framework within which to monitor and measure the changes expected to result from donor programmes and projects. Key changes described in the results chain are translated into targets and associated indicators for tracking results. Therefore, the selection of indicators is critical for results-based management systems.

The need to show results in aid for trade is growing, particularly in light of the significant additional resources directed towards trade-related activities in recent years. Increasing numbers of donors and partner countries now regularly monitor the potential impact of aid for trade and are adopting results-oriented approaches. Despite this progress, the development of a genuine performance culture based on results remains challenging.

A major challenge is designing effective results chains that connect individual project objectives with more strategic, long-term development outcomes. When it comes to measuring results at the outcome level, donors and partners alike are often confronted with the problem of attribution, *i.e.* what part of the observed changes resulted from aid-for-trade activities at the programme or project level? However, despite the unresolved question of attribution, this chapter argues that measuring results at the outcome level is essential in order to monitor and show progress towards the goals of the Aid for Trade Initiative.

As targets and results are specific to individual programmes and projects, their associated indicators also vary. Still, there are (or should be) commonalities between sector or macro level outcomes that can be quantified and aggregated into summary indices for benchmarking and cross-country comparison. A large number of indicators related to aid for trade have been generated in recent years. This chapter illustrates some of the existing indicators and presents the rationale for, and benefit of, a more aligned

approach to aid-for-trade results measurement. More specifically, it makes the case for establishing a small number of aid-for-trade indicators to enable practitioners to systematically aggregate results data across programmes and projects at the country, regional and global levels. Such enhanced transparency, in turn, will contribute to a broader effort to make aid for trade more effective.

4.2 Managing for development results

As bilateral and multilateral donors increasingly put in place essential building blocks to ensure that their activities achieve the intended objectives and targets, they are managing *for* results through articulating a chain of results – from project inputs, to activities, outputs, outcomes and long-term impacts. Results chains provide a framework within which to monitor and measure changes that are expected to result from donor programmes (Chapter 1, Section 1.6). Key changes described in a results chain are translated into targets and associated indicators for tracking results. This chapter discusses how the growing focus on performance management in development co-operation has led to managing for development results (MfDR) and examines the relevance of MfDR to aid for trade. In particular, it looks at the roles and types of indicators for measuring aid-for-trade results and explains the rationale for (and the benefit of) establishing and integrating a small set of common indicators into all programmes and projects to allow the measurement of aid-for-trade outcomes at the country level.

Aid for trade aims to "enable developing countries, particularly LDCs, to use trade more effectively to promote growth, development and poverty reduction and to achieve their development objectives, including the Millennium Development Goals (MDGs)".[32] To achieve these objectives, aid for trade (like any development co-operation programme or project that cuts across various sectors) involves complex relationships among partner country governments, bilateral donors, multilateral and regional agencies, the private sector and other NGOs. Each of these stakeholders has different priorities, operating arrangements, timeframes and financial and human resources. Therefore, increasing the effectiveness of aid for trade requires comprehensive and rigorous implementation of the aid effectiveness principles of the Paris Declaration. The Paris Declaration also stresses that partner countries and donors are mutually accountable for development results.

Recent changes in the global landscape of development assistance have led to a greater focus on transparency in and accountability for the use of development resources. Improved accountability is widely seen as an effective way to establish incentives to help strengthen local ownership and achieve results. This growing focus on development results has made "managing for results" central to the entire aid effectiveness agenda. Managing for development results (MfDR) provides a common performance management framework for achieving goals. It emphasises the importance of reviewing progress towards results, learning from what does and what does not work, and altering the overall plan if necessary. Through the Paris Declaration and the Accra Agenda for Action, development partners have committed to manage and implement aid in a way that focuses on development outcomes and impacts (rather than on process) and uses performance information to improve decision making. While it is important to get the process right, best practice in process does not guarantee tangible and meaningful results on the ground.

Against this backdrop, OECD Development Assistance Committee (DAC) donors and international agencies are increasingly putting results-based management frameworks

in place to ensure that their activities achieve the desired objectives and targets. They are going beyond process and input indicators to measure outcomes and impacts. This implies articulating a chain of results from project inputs, to activities, outputs, outcomes and long-term impacts. The results chain provides a framework within which to monitor and measure expected changes that will result from project activities. Key changes described in the results chain are translated into targets, and indicators are identified for tracking results at each step in the programme logic. Indicators are therefore a critical component of the results-based management systems.

As targets and results are specific to individual programmes, projects and country contexts, partners and donors design specific results frameworks to measure results. The types of indicators used will vary among donors and among programmes and projects. However, at the sector level there are (or should be) commonalities among outcomes that can be quantified and aggregated into summary indices for benchmarking and cross-country comparison. Thus, the unit of analysis is not a single project but a whole country programme that includes many activities implemented by different donor agencies. By monitoring the direction of these key metrics over time, donors and partners can assess the impact of their combined efforts at the sector and country level.

In some aid-for-trade sectors, such as building productive capacities, donors are already pursuing this approach by introducing sets of "universal" indicators to determine the levels of achievements and allow comparisons across countries. As more donors look to develop results frameworks for their respective aid-for-trade programmes or projects, donors and partner countries should work together to develop a manageable number of indicators in order to avoid *"an indicator cloud"* descending upon partner countries and creating *"a fog of confusion"*.[33]

This chapter explores ways to enhance performance management of aid for trade by adopting a system of managing for results. It highlights the importance of demonstrating results and impacts through MfDR. In particular, it suggests potential benefits of harmonising different results measurement systems in aid for trade. The aim is to arrive at a performance management system that would be considered feasible and sustainable. It would ideally be embedded in the results framework of all those programmes and projects which fall under the umbrella of the Aid for Trade Initiative.

4.3 How to manage for results

The need for results is recognised as a key aspect of (and prerequisite for) improved aid effectiveness. Results are those changes that can be attributed to a development measure (GTZ, 2008). More effective aid means generating more results. Governments and agencies in both donor and partner countries have become increasingly cognisant of the growing importance of accountability and performance management in development co-operation. Accountability is essential in the political arena. In response to growing calls for more efficient and effective use of development resources aimed at poverty reduction and policy reform, combined with mounting requirements for accountability, public sector agencies and development institutions have gradually introduced a range of tools to assess performance and manage for results.

The Millennium Development Goals (MDGs) adopted in 2000 encapsulated the global consensus that emerged during the preceding decade. The MDGs embody the results-based approach to international development. They contain a set of goals and measurable targets, with specific dates for achievement and performance indicators to

gauge their progress. The MDGs "represented a new departure in international thinking about how to encourage the progress of developing countries, and indeed of societies at large, and was one to which all countries, both developed and developing, committed themselves" (Manning, 2009).

A series of international events and meetings following the adoption of the MDGs were also pivotal in advocating the importance of results. Most notably, the 2005 High-Level Forum on Aid Effectiveness in Paris placed results management in development co-operation firmly at the centre of the global aid effectiveness debate. Partner countries and multilateral and bilateral donors all committed themselves to improve their management of resources and focus on the actual outcome and impact of their activities (*i.e.* development results), rather than on the inputs used or physical outputs produced, guided by mutual accountability. Specifically, in the Paris Declaration they adopted a set of actions to strengthen how they managed for development results. Subsequently, the Accra Agenda for Action (adopted at the September 2008 Third High Level Forum on Aid Effectiveness in Accra, Ghana) cited achieving increased accountability for development results as one of the key requirements for delivering on the aid effectiveness agenda. Accordingly, development partners are under increasing pressure to demonstrate results and to be accountable to their constituencies as well as to each other (mutual accountability).

What is managing for development results?

Managing for development results (MfDR) entails tracking progress and making decisions on the basis of solid evidence in the pursuit of enduring development results and impacts. Whereas conventional results-based management approaches mostly focus only on accountability, MfDR goes further, underpinning and cutting across the Paris Declaration's key pillars of ownership, alignment and harmonisation. MfDR as a concept centres on holding all development partners accountable for delivering development results (*i.e.* the outcome or impact of a development intervention) to the constituencies they seek to assist. Built on the notions of goal-orientedness, causality and continuous improvement, it is guided by the following five core principles, which reflect a broad consensus on what constitutes sound performance management:[34]

- Principle 1: Focus the dialogue on results at all phases of the development process, from strategic planning through implementation to completion and beyond. In managing for results, it is important to have a coherent approach: (i) ex ante, at the strategy and planning phase, when expected results are articulated and their likely costs and expected impact on poverty reduction and development are analysed; (ii) during programme/project implementation, when monitoring is needed to assess progress and identify necessary midcourse corrections; (iii) ex post, upon completion, when the results are assessed against objectives and other factors; and (iv) when sufficient time has passed to be able to assess sustainability.

- *Principle 2: Align actual programming, monitoring and evaluation activities with the agreed expected results.* When partner countries, development agencies and other stakeholders focus on expected results and associated results indicators, they can better align actual programming (including financial support), monitoring and evaluation activities with agreed results objectives. Partner country priorities and constraints must remain the starting point for development agencies' support strategies, and the

development agencies' planned operations, analytic support and technical assistance must be consistent with the partner country's sound development strategy.

- *Principle 3: Keep the results reporting system as simple, cost-effective and user-friendly as possible.* To the extent possible, the indicator framework for managing for results should (i) be simple; (ii) rely on country systems, supporting capacity building to the maximum extent; (iii) be geared to learning as well as accountability functions; and (iv) be harmonised in order to minimise system transaction costs and facilitate comparative analysis. The partner country and development agencies should consult on a short list of key indicators, preferably from a standardised list, for monitoring progress and assessing achievement of results. It is important to consider the chain of expected results. Managing for results aims at improved efficiency. Therefore, in choosing indicators it is essential to be selective (and not try to measure everything) and realistic (in terms of feasibility and cost). The results reporting system should remain pragmatic: start with whatever baseline data are available, including proxies; use meaningful qualitative indicators to complement quantitative indicators, or to compensate if quantitative indicators are not available; and include support for cost-efficient measures to improve data availability and country or project monitoring systems. The end goal should be a sound results-based management system that includes specific, quantifiable indicators connected to a timeline with baseline data and periodic assessments of programme and project performance against defined targets.

- *Principle 4: Manage* for, *not* by, *results by arranging resources to achieve outcomes.* Managing for results involves a change in mindset – *from* starting with the planned inputs and actions and then analysing their likely outcomes and impacts *to* focusing on the desired outcomes and impacts (*e.g.* on trade expansion) and then identifying the inputs and actions needed to get there. It also involves establishing baselines and identifying upfront performance targets and indicators for assessing progress during implementation and progress on programme or project completion. Missing key targets should be a signal for partners to analyse together whether/why things have gone off track and how they could be brought back on track, if necessary. It should not be a trigger for the rigid application of penalty rules.

- *Principle 5: Use results information for management learning and decision-making, as well as for reporting and accountability.* Information on results should be publicly available. While one of the goals of managing for results is to use results monitoring information for reporting and accountability (in the case of both partner countries and development agencies), this may potentially prompt behaviours that are overly risk-averse. Two approaches can mitigate this possibility: (i) using reports on results for management learning and decision making, taking into account lessons for better future action; and (ii) when using reports for accountability purposes, setting performance measures that reflect the level of responsibility of the actor (whether a country, development agency, ministry, institution, NGO or other stakeholder) and results that the actor can reasonably achieve. This approach recognises that even with good performance in managing for results, external factors may hinder the achievement of expected outcomes.

MfDR is a management approach that involves practical tools for strategic planning, risk management, progress monitoring and outcome evaluation. Table 4.1 lists examples of tools used to measure progress towards outcomes, report on them, and use the lessons

learned to continuously improve performance. In partner countries and donor agencies, MfDR delineates a shift from focusing on inputs and immediate outputs to the performance and achievement of demonstrable results and long-term impacts. It requires partner countries and donors to explicitly state the basis for their assessments and how decisions were made. It also entails knowledge sharing and, for joint learning in particular, the sharing of transferable lessons.

Table 4.1. Examples of tools used to manage for results

MfDR principle	Examples of tools used	Why these are important
Focus the dialogue on results at all phases of the development process	• Sector development or policy reform frameworks • Project results frameworks • Multi-stakeholder planning workshops • Inter-agency co-ordination mechanisms • Logic models (integrated in all of the above)	Results-based tools are used jointly by development agencies, and partner countries align donor support for intermediary results with national development outcomes during the planning process. Results-based tools act as reference points for ongoing implementation and measurement.
Align programming, monitoring and evaluation with results	• Annual work plans and budgets • Financial management systems	Results-based operational plans, budgets and financial mechanisms at the sector or project level describe clearly how inputs will support intermediary results leading to country outcomes.
Keep results measurement and reporting as simple, cost-effective and user-friendly as possible	• Sector-wide and/or project monitoring and evaluation (M&E) systems, including management information systems (MIS) • Sector-wide and/or project M&E operational plans and guides • Performance measurement frameworks • Sector-wide performance monitoring strategy • Annual quality control reviews for service delivery to clients/beneficiaries • Data source assessment/review	M&E systems, plans, frameworks and instruments describe the indicators for intermediary results at the sector and project levels, describe methods for data collection and analysis, assign M&E roles and responsibilities, and provide standardised methods for assessing progress.
Manage for, not by, results by arranging resources to achieve outcomes	• Special studies (thematic or value-for-money) and policy reviews • External and internal monitoring reports • Mid-term social impact assessments and/or sector/thematic outcome evaluations • Technical milestones linked to financial disbursement schedules • Sector programme reviews • Performance and financial audits • "Scoreboards" and periodic activity reports	Studies, reviews, assessments and monitoring all investigate issues related to results achievement and suggest means of adjusting implementation strategies as required, at either the sector-wide or project levels.
Use results information for learning and decision making as well as reporting and accountability	• Annual sector-wide or project performance reports • Stakeholder consultations • Analysis of evaluations	Reports and consultations provide government officials, sector ministries, development agencies, civil society, grassroots beneficiaries and other key stakeholders with performance information on progress towards intermediary results and country outcomes at the sector and project level.

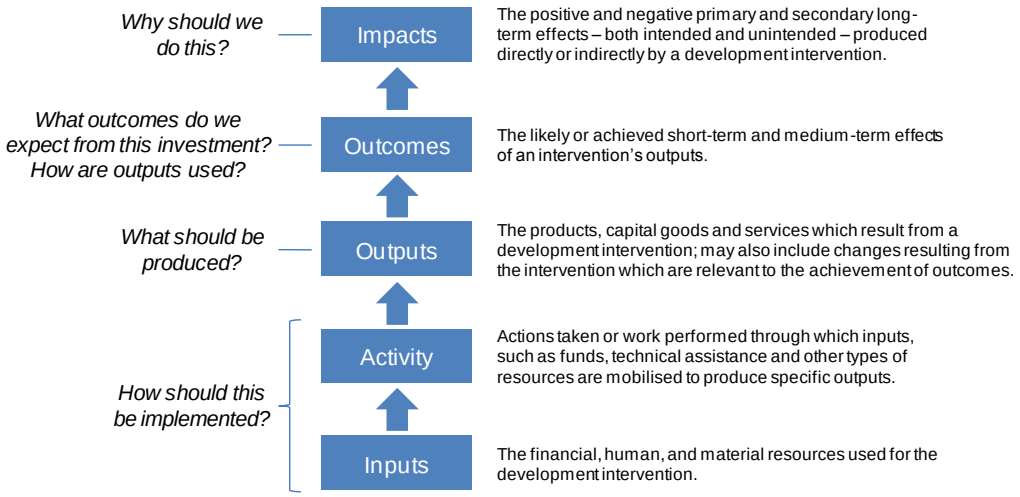
Source: "MfDR Principles in Action: Sourcebook on Emerging Good Practices" (first edition) (*www.mfdr.org/Sourcebook/1stEdition/MfDRSourcebook-Feb-16-2006.pdf*)

How to manage for development results

Periodic measurement of results provides the basis for correction and adjustment. It also enables governments and donor agencies to better guide their performance to keep programmes/projects on track and maximise their outcomes. This process is centred on a strong notion of causality. Cause-and-effect relationships between development interventions and the intended results are usually depicted using a results chain.

A results chain shows the change processes necessary to achieve desired objectives, beginning with inputs, moving through activities and outputs, and culminating in outcomes, impacts and regular feedback to the responsible officials and civil society stakeholders. Figure 4.1 is a graphical representation of how the development objective is to be achieved through a results chain, including causal relationships and underlying results questions to enable development results. Thus, inputs are used to undertake project activities that lead to the delivery of outputs (goods and services), lead to the attainment of the project purpose (outcomes), and then contribute to project goals (impact).

Figure 4.1. Results chain for results-based management



Source: OECD (2009).

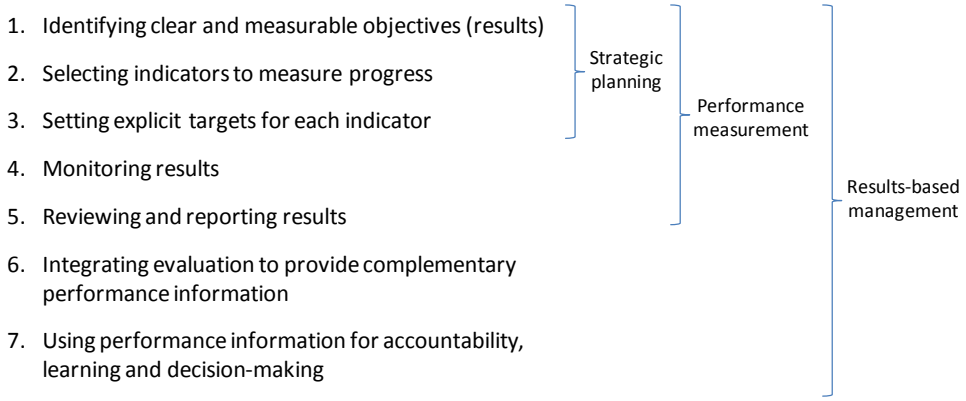
Use of the results chain, in turn, implies that an effective performance assessment can only take place if clear objectives and verifiable targets at output, outcome and impact level have been defined at the outset of the programme. It is necessary to demonstrate a chain of results from programme/project inputs, to activities, outputs, outcomes and long-term impacts. By focusing on expected results from the outset of programme/project design, it is possible to measure the associated baseline and target values and indicators for tracking the performance of each proposed investment.

For instance, the World Bank's Trade and Transport Facilitation Programme in Southeast Europe (TTFSE) successfully achieved its objective of reducing the non-tariff costs of trade and transport at selected border crossings in eight countries through, *inter alia*, setting up a transparent, comparative public performance monitoring system across the pilot sites using a set of clearly defined, standardised performance indicators (OECD/World Bank, 2006). The programme found that measuring progress using indicators was vital for successful monitoring and implementation. In order to track

results, this approach places greater emphasis on setting clear and appropriate targets, a proper baseline, mechanisms for information collection, and a fully integrated results chain to monitor development and assess what works and what does not work.

This method also lies at the core of any results-based management approach. Figure 4.2 lists the key phases or steps of results-based management.[35] The first three steps generally relate to a results-oriented planning approach ("strategic planning"). The first five, together, are usually included in the concept of "performance measurement". All seven steps combined are essential to an effective results-based management system.

Figure 4.2. The seven phases of results-based management



Source: Adapted from Binnendijk (2000).

However, an increased focus on results can lead to an overemphasis of the measurement dimension and to a multiplication of indicators, which could ultimately prove detrimental to management. That is why MfDR requires that goals be clear, measurable, limited in number and concrete, with time-bound targets. At the same time, they must be expressed in human development terms and linked to national goals (*i.e.* as development outcomes). In the case of the TTFSE, the programme design was kept flexible to facilitate minor adjustments in initial plans, paying close attention to achieving "realistic" results. The programme objectives sometimes had to be scaled down due to changes in initial conditions or when preliminary performance indicator results failed to meet the initial expectations.

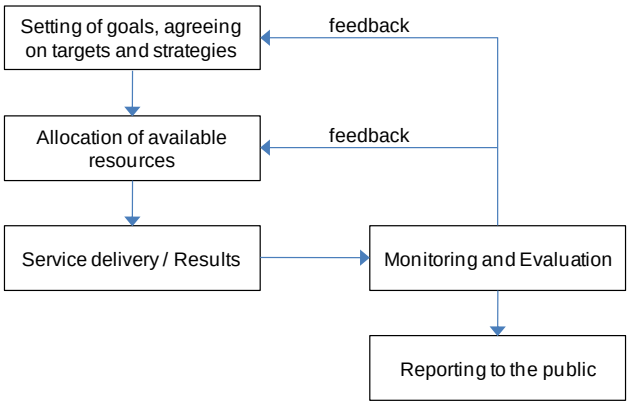
The role of evaluation

Within the MfDR framework, the performance management and reporting cycle has five core elements: (i) setting concrete goals and agreeing on targets and strategies; (ii) linking policy measures and budgets (*i.e.* performance-based budgets[36]); (iii) monitoring and evaluation (M&E); (iv) reporting on performance to the public for accountability; and (v) continuous feedback of performance information into decision making (readjustment of policy measures and resource allocation). Figure 4.3 is a simple graphical representation of the MfDR cycle.

As shown in this figure, M&E is an integral part of the MfDR cycle. Monitoring and evaluation can be viewed as complementary but distinct processes. Whereas monitoring answers the questions of what results were achieved and whether targets were met,

evaluation provides means of learning about why and how those results were achieved. Together M&E provides, in real time, important sources of performance information that can contribute to management learning (enabling managers to make timely changes to improve performance if things are off track) and building a knowledge base to improve development policies and programming. Effective reporting promotes transparency and mutual accountability, which in turn support good governance.

Figure 4.3. The MfDR cycle



Source: OECD Secretariat

Because of timing, as well as the need for more in-depth analysis, some performance issues, such as long-term impact, attribution, cost-effectiveness and sustainability, can probably be better addressed by evaluation than by routine performance reports. Evaluation is critical in looking beyond the implementation process (*i.e.* whether activities were implemented in a timely manner and outputs were obtained) to cover the extent to which activities and outputs contribute to reaching the desired outcomes and impacts (including relevance, attribution, cost-effectiveness, sustainability and unintended results).[37]

Institutionalising MfDR at the national and regional level

Successful application of MfDR requires a shift in mindset and focus. It needs to be based on a leadership and management culture that is focused on achieving results (and not just on compliance) and anchored in solid evidence for decision making. MfDR also needs to be a country-led process so that it promotes, rather than weakening, local structures of accountability and leads to a sustainable improvement in development outcomes. In recognition of these challenges, the 2005 Paris Declaration called for strengthening country capacities and demand for results-based management and set in motion a process to achieve this.

Within that framework, partner countries maintain ownership, define their priorities, agree on expected results through a participatory process, and manage effective relationships with donors based on their contributions to national goals (public accountability). Donors can, in turn, strengthen the impact of their efforts by aligning them with local strategies, and by monitoring and assessing their contributions to outcomes to ensure greater aid effectiveness. This is why it is important for all stakeholders to sign up to the shared goals and strategies that have been agreed upon, and

to keep those responsible for implementing the process accountable with regard to delivering results. Partner country governments are therefore assisted both in driving forward reforms and in remaining accountable for the results they achieve.

Among the MfDR milestones so far have been the establishment of region-wide communities of practice (CoPs) on MfDR in Africa, Asia, and Latin America and the Caribbean to exchange and disseminate MfDR knowledge and good practices among practitioners within and between regions.[38] Steps are being taken in each region to institutionalise MfDR at the national and sub-national level through the establishment of national and local CoPs.

A diagnostic tool, the MfDR Capacity Scan (CAP-Scan), has also been developed as part of the OECD/DAC Global Partnership on MfDR and is supported by a working group of donors.[39] CAP-Scan allows partner countries to conduct assessments that provide a clear view of strengths and capacity gaps, develop actions to address resource needs and target donor support. Country-owned development is the cornerstone of the Paris Declaration.

4.4 Managing aid to achieve trade results

The self-assessment results of the 2009 Aid for Trade Initiative monitoring survey indicate that more partner countries are accepting the tenets of mutual accountability and results-based management, while donors are responding by increasingly co-ordinating and aligning their aid-for-trade efforts. Moreover, the same survey found that most partner countries have put in place (or are working towards setting up) mechanisms to discuss the results of trade-related programmes (OECD/WTO, 2009).

Two-thirds of partner countries reported that they regularly monitored and evaluated their trade-related programmes, and that they frequently used donors or joint donor-partner country arrangements (including sector-wide approaches). Low-income countries tended to have higher rates of monitoring and evaluation than middle-income ones. While 9 out of 28 Least Developed Countries (LDCs) responding reported that they rarely or never engaged in monitoring, their situation should improve once the Enhanced Integrated Framework's new monitoring and evaluation framework becomes operational. Furthermore, almost all partner countries regularly took part in dialogues with relevant stakeholders about the formulation and implementation of their trade strategies. In terms of implementation, partner countries underscored the need for strengthened capacity building and improved ownership, which play a key role in determining the effectiveness (and thus the results) of aid for trade (OECD/WTO, 2009).

This aid-for-trade monitoring survey also revealed that more donor agencies were adopting results-oriented approaches (mostly at project level).[40] According to the survey results, for instance, over 40% of bilateral and multilateral donors responding reported that they monitored the potential trade impact of their aid programmes/projects (OECD/WTO, 2009).[41] Moreover, many donors were taking steps to make their respective monitoring and evaluation frameworks more results-oriented. Finland and Sweden highlighted their plans to strengthen their monitoring and evaluation systems by developing specific indicators, which would include cross-cutting themes such as gender. Germany reported that it was looking at ways to design a mechanism, including impact chains and indicators, to monitor the implementation of environmental performance assessments (EPAs). Switzerland, in co-operation with the United Nations Industrial Development Organization (UNIDO), was developing a standard logical framework for

each of its activity categories (*e.g.* export promotion, competition and consumer protection policy) as part of its efforts to strengthen results-based management. All of these donor efforts are aligned with and contribute to the WTO Aid for Trade Task Force's emphasis on "the need for concrete and visible results on the ground".

Challenges: from principles to practice

Despite all this momentum and apparently widespread acceptance, the development of a genuine performance culture based on results remains a challenge, not only for partner country governments but also for donor agencies. The results of the 2008 Paris Declaration monitoring survey also indicate that while some progress had been made with regard to the number of countries establishing sound, results-based monitoring frameworks,[42] the pace of progress was still too slow (OECD, 2008a).

Disincentives to report failures are part of the explanation.. Reporting failures may generate the perverse result that fewer funds flow to a specific country or sector, although the reported failure might serve to enhance the effectiveness of the next round of investment. A new study assessing the results of reporting practices in donor agencies reveals that many of the key MfDR elements (*e.g.* baselines, targets, results chains) remain missing in donor reporting.[43]

There is still a need to raise awareness of the importance of "results" monitoring, both to assess the impact of aid for trade and to justify continued support for the Aid for Trade Initiative (OECD/WTO, 2009). A major challenge is how to design effective intervention logics or results chains that connect individual programme or project objectives with more strategic long-term development outcomes and impacts. This challenge is often associated with attribution problems, a significant time lapse (between the design of the intervention, its implementation and its impact), lack of credible data, and difficulties in assessing often intangible capacities.

In the case of aid for trade, these problems are compounded by the wide scope and multiple objectives of the Initiative. It is difficult to identify clear and verifiable indicators for measuring the effectiveness and impact of aid-for-trade programmes and projects. Donors have indicated in their self-assessments that identifying and measuring trade-related outcomes is a real challenge (OECD/WTO, 2009). The European Commission, for instance, stated that monitoring and evaluating the trade impacts of specific aid programmes is unfeasible given the many external factors that influence trade (OECD/WTO, 2009).

To overcome these constraints, one approach would be to identify and introduce a menu of indicators for aid-for-trade interventions to allow for the aggregation of results at the country level and for cross-country comparability. The use of indicators is crucial in results reporting, as they specify how expected results have been measured and also define the data to be collected. Where data are hard to come by, proxy indicators could be used. This approach would permit countries – through harmonised results measurement and reporting practices – to observe and compare strengths, weaknesses and gaps across donors and track progress over time. This, in turn, would help to identify specific actions that could help improve the impact of aid for trade.

4.5 Using indicators to improve performance

Aid-for-trade results are difficult to measure. A range of indicators may need to be used in order to obtain a full picture. The need to show results in aid for trade is growing, particularly in light of the significant additional resources that have been directed towards trade-related activities in recent years. Stakeholders will need a clearer sense of the outcomes and impacts of aid-for-trade efforts. The ability to demonstrate results can, in turn, help to attract greater political and public support within donor and partner countries for more and better aid for trade. One of the key objectives for the way forward, as stated by WTO Director-General Pascal Lamy at the Second Global Review of Aid for Trade in July 2009, was the continuation of evaluation work with a specific focus on evaluating the "impact" of aid for trade. Some work has already been undertaken to investigate ways of identifying indicators to measure the impact of aid for trade (Box 4.1).

Box 4.1 Using indicators: some evidence from regression analyses

A number of recent studies have used World Bank indicators to show that aid-for-trade facilitation has a significant cost-reducing effect with regard to the costs of handling exports. Using a cross-section gravity model, Hoekman and Nicita (2010) have demonstrated that domestic trade costs are both a statistically and quantitatively significant determinant of trade volume (a more limiting factor for international trade than tariffs). Lack of trade facilitation and related infrastructure (i.e. high trade costs) substantially reduces trade volumes. Improving the Logistics Performance Index (LPI) of low-income countries to the level observed in high-income ones would increase their trade flows by more than 50%, everything else being equal. Similar results were obtained for the effect of internal trade costs as captured by the Trading Across Borders indicators for exports and imports: a 10% reduction in the costs associated with importing (or exporting) would increase imports (or exports) by about 5% (Hoekman and Nicita, 2010).

Calì and te Velde (2009) used the Trading Across Borders indicators to estimate whether facilitating aid for trade had any impact on trade costs. In terms of costs to export and import a standard size (20-foot) container, they estimated that a USD 1 million increase in aid-for-trade facilitation would reduce the "per container" cost of packing and loading the container, transporting it to the port of departure and unloading it onto a vessel or truck by 6%, or about USD 70. The return on aid-for-trade facilitation (aggregated savings) could be substantial: more than 7 million 20-foot containers were loaded and unloaded in African ports alone in 2000 (Calì and te Velde, 2009).

In the same study, Calì and te Velde measured the impact of aid for trade by matching sub-sets of aid with more specific sector-level changes (or outcome variables). They analysed aid to different sectors (food production, manufacturing, mineral extraction and tourism) and then related sectoral aid to sector-specific exports to determine whether sectors in a country that received more aid for trade experienced relatively faster growth in their exports, or whether a sector's exports grew faster in years when that sector received relatively higher levels of aid for trade. However, their analysis includes a caveat: it is possible that donors will allocate more aid for trade to countries that are relatively better performing and/or faster reforming, biasing the impact of that aid (*i.e.* the endogeneity of aid for trade). Indeed, some recent studies (Brenton and von Uexküll, 2009; Calì and te Velde, 2009) have found that the positive effects of sector- or product-specific aid for trade (building productive capacity) on exports appear to coincide with aid allocation skewed towards already well performing sectors.

In terms of determining the indicators that are most relevant, Gamberoni and Newfarmer (2009) identified a set of indicators to measure "potential demand" for aid for trade. They looked at the determinants of a country's international competitiveness and the degree to which it was integrated into the global economy. In particular, they focused on three key determinants of trade performance that governments could influence through their trade and trade-related policies and the investment of aid-for-trade resources. These were: (i) a good regulatory environment; (ii) good trade facilitation policies and practices; and (iii) a good base of trade capacity (in particular, trade-related infrastructure). Gamberoni and Newfarmer identified five key indicators

of trade performance. They then looked at the causes of poor trade performance, which they defined as trade-related domestic capacity constraints (*i.e.* infrastructure, institutions, and policy-induced price incentives), and identified corresponding indicators. Table 4.2 lists the resulting sets of possible macro-level aid-for-trade indicators to monitor trade capacity and trade performance that are readily available.

Table 4.2 Macro-level aid-for-trade indicators for monitoring trade capacity and trade performance

Dimension	Indicator	Source
Trade performance	Real growth of exports of goods and services	(World Bank) (WB) World Trade Indicator
	Change in export market share of goods and services	WB, World Trade Indicator
	Competitiveness effect (change in market share)	International Trade Centre (ITC), Trade Performance Indicator
	Demand effect (change in market share)	ITC, Trade Performance Indicator
	Index of export concentration (Herfindhal)	WB, World Trade Indicator
Infrastructure	Quality of transport and information technology (IT)	WB, Logistics Performance Index
Institutions	Efficiency of customs	WB, Logistics Performance Index
	Time to export/import	WB, Doing Business
Incentives	Trade restrictiveness index (tariffs only)	WB, World Trade Indicator
	Share of tariff lines with domestic peaks	WB, World Trade Indicator

Source : Adapted from Gamberoni and Newfarmer (2009).

Attribution or contribution?

Managing for results is about delivering aid in a way that focuses on development outcomes and impacts. However, the further one moves from specific project- or activity-level results to indirect results, the more difficult it becomes to attribute those results to a specific intervention. As one moves along the results chain, a multitude of external factors interact with and influence each element. A few outputs may also result in several outcomes, and in even more impacts. This is why it becomes more difficult to attribute causes to the final impacts. Donors are often confronted with the problem of the "missing middle" (*i.e.* what part of the observed changes resulted from aid-for-trade activities at the project output level?). These indirect results depend on the interplay between many different factors and actors that cannot be influenced by the programme or project (GTZ, 2008). There are several ways to estimate attribution, which vary in their level of sophistication (Table 4.3). Nevertheless, the complexity of assessing the impact of individual donor projects on the beneficiary's overall trade capacities and performance appear to be a key methodological challenge identified by most evaluators (OECD, 2007).

In the case of aid for trade, this problem is compounded by the number of other variables that may also affect a country's trade performance, such as geographical characteristics, the legal system, regional effects, income levels, population size and governance (Calì and te Velde, 2009). For this reason, it would be impossible to track all of the causal factors that affect the attainment of the higher-level results. Rather, as Toffolon-Weiss *et al*. (1999) explain, "the results-framework approach focuses on key results that can be influenced by the intervention and will contribute to the desired outcome." Moreover, Elliot (2007) argues that "[…] as long as measures of ultimate development objectives are moving in the desired direction, then donors should not be overly concerned about being able to attribute the results to aid. If the measures are not showing improvement, then closer scrutiny of the effectiveness of aid delivery and implementation, relative to other factors that affect growth, development and poverty is merited."

Conducting joint evaluations can also help alleviate the administrative burden on partner countries and determine the collective impact of donors' efforts (especially *vis-à-vis* new modalities such as general budget support and sector-wide approaches), overcoming to some degree the attribution problem individual donors and agencies face (OECD, 2007).

Therefore, despite the methodological challenge, monitoring the indirect results and collective contributions of development partners is important to ensure that progress is being made on the ultimate objective of the Aid for Trade Initiative. This requires, among others, results-based indicators for continuous monitoring and evaluation, which the rest of this chapter will now focus on.

Table 4.3. Methods of estimating attribution

Method	Application	Advantages	Disadvantages
Opinions of key informants and expert interviews	May be important when the key change is driven by one person (*e.g.* a politician changing a policy)	Low cost	May be influenced by interviewer; likely to be somewhat subjective.
Comparison of treatment and control group (randomised samples)	When samples are large enough, in measuring changes attributable to one step in the results chain (probably not feasible for the whole model in one trial)	Considered by statisticians to be the most reliable way to measure results (albeit based mainly on experience with simple/single treatments)	Difficult to design and administer if the treatment group is self-selecting (*e.g.* buying a service). In that case, a randomised sample would need to be refused a service they tried to purchase.
Quasi-experimental design (differences of differences: comparing before and after for treatment and control groups)	Often appropriate for pilot efforts and/or measuring attributable changes with regard to one step in the results chain	More approximate, in acknowledging that the control group is not an exact control	Cheaper than randomised controlled trials, but still expensive. Careful design and measurement needed to ensure accuracy. Not valid when the target group is unique, as is often the case with large urban clusters, or when interventions can influence the control group as well as the treatment group.
Participatory approaches (focus groups, etc.)	Where the change in behaviour might have been caused by different factors	May be the only way to show attribution in some cases	May be subjective, open to bias (*e.g.* high subsidies may attract positive ratings, even though not sustainable).
Observation	Where attribution is fairly clear (*e.g.* resulting from new technology)	Low cost	May not be perceived as convincing, especially where attribution is not obvious.
Regression analysis	Where a wide range of data can be accurately gathered	Can be reasonably accurate if well designed and executed	High level of skill needed; accuracy relies on identifying and gathering data on other significant factors contributing to the change.
Extrapolation of attribution proven in pilot or case study	Where funds are not available for large-scale measurement	Low cost, relatively convincing	Needs periodic verification by other means (*e.g.* through surveys or additional case studies).
Trend analysis	Where other, large trends are very significant and trends can be reasonably tracked and estimated	Takes into account larger economic and market trends; relatively low cost	Risks assuming that the identified and measured trends are the only (or main) ones applicable; best used, therefore, in combination with other methods.
Case studies analysing behaviour and performance changes at each step of the results chain	Where qualitative understanding is needed to interpret quantitative data	Low cost; can be a good indication of attribution if well designed and executed	May not represent the universe of beneficiaries; can be time consuming; may be influenced by interviews.

Source: Adapted from DCED (2010).

Typology of indicators

What actions are expected to lead to what results? To establish the basis for measuring impact, relevant programme or project indicators are needed for each step in the results chain (*i.e.* inputs → activity → outputs → outcomes → impacts) in order to show how changes at each level lead to changes at the next, ultimately impacting on the long-term development objectives (*e.g.* poverty reduction). These indicators comprise *monitoring indicators*, which primarily relate to inputs and outputs, including progress in the use of allocated funds, and *performance indicators*, which primarily focus on outcomes and impacts.[44]

Indicators provide indications (or yardsticks) of the extent of changes and the achievement of objectives. Some of the basic definitions of indicators used by some donors to ensure clarity are shown in Table 4.4.[45]

Table 4.4. Definitions of indicators used by some donors

	INPUTS	ACTIVITY	OUTPUTS	OUTCOMES	IMPACTS
EU	**Input indicators:** Financial, human, material, organisation or regulatory resources mobilised during the implementation of the intervention.	**Activity indicators:** Implementation and management process.	**Output indicators:** Goods and services that are delivered under the responsibility of the managers of the intervention.	**Results indicators:** Immediate effects of the intervention for its direct addressees.	**Impact indicators:** Far reaching and indirect consequences of the intervention.
MCC		**Process Milestone Indicators** measure progress towards the completion of project activities; a precursor to the achievement of output indicators and a way to ensure the work plan is proceeding on time to sufficiently guarantee that outcomes will be met as planned.	**Output indicators** directly measure project activities. They describe and quantify the goods and services produced directly by the implementation of an activity.	**Outcome indicators** and **Objective indicators** measure the intermediate (medium- to long-term) effects of an activity or set of activities and are directly related to the output indicators.	**Goal indicators** measure the economic growth and poverty reduction changes that occur during or after implementation of the programme. For MCC Compacts, goal indicators will almost always be a direct measure of income and/or poverty.
JICA	**Input indicators** express the resources introduced into the project.	**Process indicators** express the progress of activities within the project, e.g. indicators showing how much a budget has been depleted.	**Output indicators** measure the results of the project activities, or the products (goods & services) created by the project, i.e. the outputs of the logic model.	**Direct Outcome indicators** express direct and short-term changes brought about by the project. **Intermediate Outcome indicators** express medium-term changes brought about the by project for beneficiaries and society.	**Final Outcome indicators** express broad-based and long-term changes brought about by the project for beneficiaries and society (also referred to as impact indicators).

Sources: Based on European Commission (2006), MCC (Millennium Challenge Corporation) (2009) and JICA (Japan International Cooperation Agency) (2005).

Although donors may use slightly different terms and/or definitions (Table 4.4), indicators are generally classified into three types:

Input, activity and output indicators

Micro-level (*i.e.* input, activity and output) indicators concern narrow, project-related measures. They are used for project-level management and help to track progress in implementation based on benchmarks and baseline data. *Input and process indicators* provide information on financial, human, material, organisational and regulatory resources mobilised during the implementation of the intervention. *Output indicators* describe and quantify goods and services produced directly by the implementation of an activity. These results, in turn, represent the building blocks that provide the basis for

(and valuable monitoring of information concerning progress towards) attainment of the desired macro-level (*i.e.* outcome and impact) results.

Outcome indicators

Country-level results consist largely in the targets/objectives achieved by individual projects. Outcome indicators are used to measure the "intermediate" effects of an activity or set of activities targeted by aid for trade. They are directly related to the output indicators. Outcome indicators refer to the degree to which results are achieved over time, and so can be further classified as short- and medium-term outcomes.

Short-term outcome indicators are more directly linked (*i.e.* do not need additional intermediate results to understand the linkage) with the short-term changes brought about by the project outputs. They are a set of "lower-level" intermediate results that need to be reached to achieve a longer-term objective. Unlike estimating a relationship between aid and growth, the links between aid for trade and trade-related performance (enhanced trade capacity) are more direct. For instance, in terms of aid-for-trade facilitation, improvements in trade costs and competitiveness may be closely linked to the aid provided. The results at this level contribute to further changes in the medium to long term that can no longer be causally attributed to an individual measure.

Medium-term outcome indicators are used to measure medium-term changes brought about (indirectly) by the programme or project on beneficiaries. An intermediate result is a discrete result or outcome essential to the achievement of the final outcome (or another intermediate result critical to achieving the final outcome). For instance, changes in trade or investment flows are the expected result of aid for trade (*e.g.* transport infrastructure) at this level of objectives. However, these occur for any number of reasons and lags between the provision of aid and improved trade performance can be long, making attribution of this result to aid difficult. Results at this level cannot be attributed to a development measure, but its outputs do plausibly contribute to achieving the objectives.

It is at the level of outcomes that there is potential to adopt a focused set of standard indicators at the sector level which would help aggregate results data from across aid-for-trade projects at the national, regional and global level. The aggregation of data is important in reporting results at a macro and organisational level. This set of indicators will capture the development partners' "collective contributions" to results achieved on the ground.

Impact indicators

Impact indicators are used to measure broad-based, long-term changes (directly or indirectly, intended or unintended) for the beneficiaries (*i.e.* ultimate goals, growth and poverty reduction) brought about by a programme or project. These changes are more difficult to predict and to attribute to the results chain of causation. They lie beyond the "attribution gap" (GTZ, 2008). It may be recalled that in the case of the Aid for Trade Initiative the goals are to "enable developing countries, particularly LDCs, to use trade more effectively to promote growth, development and poverty reduction and to achieve their development objectives, including the Millennium Development Goals (MDGs)" (WTO, 2006). Given such broad-based goals, it will be even harder to clearly trace outcomes and impacts back to micro-level aid-for-trade activities, let alone establish the counterfactual. Instead, this level of results tracks development progress at the national, regional and global level through selected macro-level outcomes to which the Aid for

Trade Initiative contributes (*e.g.* growth and income distribution effects, as well as contribution to poverty reduction). These outcomes result from collective and collaborative action by development partners over the long term. They represent the major development outcomes the Aid for Trade Initiative is seeking, and to which aid-for-trade interventions aim to *contribute* (rather than to receive direct or individual attribution).

Hence, indicators at this level will include a combination of: (i) basic *contextual indicators* (relating to a country's demographic, social and economic conditions) to provide the necessary background for understanding development concerns; and (ii) *global indicators* (relating to development goals and objectives set forth in the United Nations and other international forums) to provide further insights into thematic or cross-cutting issues of major development concern such as gender and environment. Contextual indicators are often designed to highlight the specificities of a local context and therefore do not provide a global and normative view of a country's situation.

4.6 Towards a menu of aid-for-trade indicators

Although attribution is a problem, outcome indicators are useful to point towards the direction of changes with which a programme or project can be associated. While the choice of indicators tends to be driven in part by data availability, many indicators related to different areas of aid for trade are available today (Annex B).

The key to a light but effective monitoring system is to focus clearly on results. The first step is to narrow down the variables that might affect trade performance and identify those that can be related to the investment of aid-for-trade resources. The United Nations Development Programme (UNDP) has set out a number of caveats with regard to identifying aid-for-trade indicators:[46]

- Provide factual information about implementation progress;
- Do not try to point to possible causation between different variables;
- The story emerging should be recognisable and conducive to a constructive dialogue focused on where further improvements are required;
- Do not aim to be comprehensive;
- Indicators should remain a political tool for assessing overall trends and progress;
- Presentation of results should be based as much as possible on benchmarking and cross-country comparisons.

As more donors start to develop results frameworks for their aid-for-trade programmes,[47] introducing a limited number of indicators to measure results would enable both donors and partners to "add up" these results across programmes and projects for benchmarking and for cross-country comparability. This would also be consistent with the spirit of the Paris Declaration and the Accra Agenda for Action, which called for the maximum effort to arrive at harmonised international indicators.

The case for a menu of indicators

As already pointed out, the need to show results in aid for trade is growing, particularly in view of the significant additional resources directed towards trade-related

activities in recent years (OECD/WTO, 2009). However, a plethora of indicators are being generated and currently used by various partners to assess progress towards achieving specific aid-for-trade results. For instance, in its 2001 guidelines, *Strengthening Trade Capacity for Development*, the DAC identified a number of indicators for assessing donor support for trade capacity building (OECD, 2001).[48] In OECD (2007) there is a list of performance indicators used by various donor agencies to assess different types of trade policy and regulations and trade development projects. An alternative framework was presented by Elliot (2007), who proposed a number of potential indicators at the project output level, as well as the outcomes and the impact levels, to assess the impact of donor support aimed at addressing supply-side constraints (Table 4.5).

Table 4.5. Possible indicators of supply-side constraints

Achievement of immediate project goals (outputs)	Measures of trade costs and competitiveness (short-term outcomes)	Measures of trade and investment flows (medium-term outcomes)	Impact on ultimate goals
• Kilometres of roads built and maintained • Increases in sea/air port capacity • Increases in access to landlines, cell phones, internet • Access to credit • Reductions in power outages • Access to cold storage, especially in rural areas • Increased compliance with SPS and other international standards • Rationalisation, harmonisation of regulations related to trade, transit in regional trade agreements, especially involving land-locked countries	• Reduction in number of forms required to import/export • Reduction in days for goods to clear customs • Reduction in trade taxes, especially on key technologies, other inputs • Reduction in internal transit time to market, port or end user • Reduction in total time to get goods to destination • Reduction in the share of output not reaching market due to delivery delays • Competition measured by market shares of top five or ten firms providing logistics, transportation services • Reduction in transportation costs (changes in CIF/FOB) • Size of inventories held • Effects of aid on exchange rate	• Increased capacity in sectors producing tradable goods and/or services • Increased value-added in tradable goods and/or services sectors • Increased firm-level productivity • Change in global export shared (total and in key sectors) • Diversification of exports (share of top five products in total exports) • Increased private investment (foreign or domestic) in and around infrastructure projects and in productive sectors receiving assistance	• Higher employment levels in tradable goods and/or services sector • Increased number of subsistence farmers engaging in market activities (local or export) • Lower shares in economic activity/employment for informal sector • Higher and sustained growth following increases in trade • Higher overall employment if growth stimulated • Reductions in poverty rates

CIF/FOB: Cost, insurance and freight/free on board

SPS = Sanitary and phytosanitary

Source: Adapted from Elliot (2007).

Harmonisation of indicators among development partners is neither feasible nor desirable given the differences in operational needs and strategic priorities. Micro-level indicators, for instance, are used to monitor specific programme or project needs and priorities. Thus, they vary by programme or project and cannot be aggregated across programmes or projects. Focusing on the programme or project as the unit of analysis for performance management has limitations. Often programme or project logframes lack a strategic and long-term focus and are inappropriate for dealing with country sector-wide programming modes that are jointly supported by numerous development partners. However, some common themes emerge among desired results at the sector outcome level, and such similarities can be enhanced and used.

There is considerable benefit in developing and integrating a small number of "universal" indicators across all aid-for-trade interventions. This allows benchmarking of

aggregated progress in building trade capacity at the country level. The challenge is to select indicators broad enough to capture the wide range of aid-for-trade programmes and projects while still providing credible information on how aid for trade is achieving results in trade capacity and in helping partner countries to expand trade and develop.

Several donors and donor programmes/projects in the realm of aid for trade have adopted common-indicator approaches to manage and account for results (Box 4.2). The idea behind this is to establish a *menu* of indicators, although not a definitive or comprehensive one. It should be reasonably representative of the essential characteristics of aid for trade per activity sector, as defined by the Aid for Trade Task Force (Table 4.6), but at the same time be shaped by data availability. These indicators, taken together, provide a sense of progress and challenges at the country level and contribute to a broader effort to make aid for trade more effective. Such a menu would also be subject to improvement over time as the knowledge base improves.

Box 4.2 Examples of common-indicator approaches

The monitoring framework of the Millennium Challenge Corporation (MCC) in the United States focuses on objectively measurable outcomes to reduce "the ambiguity and sometimes conflicting objectives that can undermine development assistance" (Wiebe, 2008). The MCC uses sector-wide "common" indicators (specified at all indicator levels, *i.e.* activity, output and outcome) to measure progress and aggregate results across recipient countries within certain sectors.[1]

The Development Outcome Tracking System (DOTS) of the International Finance Corporation (IFC)[2] uses standardised, industry-specific indicators to track the outputs and outcomes of IFC-supported companies and to capture overall development results (although not to estimate results specifically attributable to investments). There are also corporate-level indicators to track outcomes related to broad themes that are relevant to all projects (*e.g.* corporate governance, environmental and social improvements). Results from individual projects are aggregated for comparison and reporting within industry sectors, as well as across the IFC as a whole.

The Enhanced Integrated Framework (EIF) has defined explicit targets and corresponding indicators for key results areas at each level of the logical framework (*i.e.* short-term outcomes, medium-term outcomes and impacts).[3] These indicators are used to aggregate results across all EIF partner countries and assess the EIF's overall contribution to the performance of LDCs (Smith, 2009). Because the scope and range of each project is specific to the intended results of the project – so that indicators will also vary by project – the EIF monitoring and evaluation framework does not specify project-level output indicators.

The leading multilateral development banks (MDBs) have formed a working group on MfDR with the aim of improving results monitoring and reporting at the corporate level.[4] The group established a common performance assessment system (COMPAS) that has corporate-level development results frameworks with "shared accountability" for outputs and outcomes. MDBs are now discussing ways to develop and use a common set of standardised sector indicators to capture aggregate MDB contributions to results in selected areas (*e.g.* agriculture, SME development, road transport and energy) at country, regional and global level.

1. *www.mcc.gov//pages/activities/activity/monitoring-and-evaluation*
2. "IFC's Development Outcome Tracking System (DOTS)" (*www.ifc.org/ifcext/devresultsinvestments.nsf/Content/DOTS).*
3. "Enhanced Integrated Framework for trade-related assistance to Least Developed Countries (LDCs)" (*www.enhancedif.org/*).
4. Participating MDBs include the African Development Bank (AfDB), Asian Development Bank (ADB), European Bank for Reconstruction and Development (EBRD), International Fund for Agricultural Development (IFAD), Inter-American Development Bank (IDB), Islamic Development Bank (IDB) and World Bank Group (International Bank for Reconstruction and Development, IBRD, and International Finance Corporation, IFC). See "Managing for Development Results" (*www.mfdr.org/Compas/index.html*).

The practical application of this results framework led to the development by the OECD and the WTO of the "aid-for-trade at a glance" country fact sheet.[49] This tool is meant to enable rapid cross-country comparisons, based on a limited number of indicators

drawn from existing sources. However, the country fact sheet should be seen as an "evolving" tool. It does not provide all the answers with regard to how to carry out MfDR in aid for trade, nor does it claim to be able to do so. It has provided a starting point to help countries discuss where gaps remain in making aid for trade more results-oriented and ways to overcome them.

Table 4.6. Aid for trade by category

Sector	Sub-sector	OECD Creditor Reporting System (CRS) code
Trade policy and regulations	Trade policy and administrative management	33110
	Trade facilitation	33120
	Regional trade agreements (RTAs)	33130
	Multilateral trade negotiations	33140
	Trade education/training	33181
Economic infrastructure	Transport and storage	21010 to 21081
	Communications	22010 to 22040
	Energy supply and generation	23010 to 23082
Building productive capacities	Business and other services	25010
	Banking and financial services	24010 to 24081
	Agriculture	31110 to 31195
	Forestry	31210 to 31291
	Fishing	31310 to 31391
	Industry	32110 to 32182
	Mineral resources and mining	32210 to 32268
	Tourism	33210
Trade-related adjustment		33150

Source: OECD Creditor Reporting System (CRS) Directives (*www.oecd.org/document/56/0,3746,en_2649_34447_1948088_1_1_1_1,00.html*)

This tool should be further strengthened to improve (on a country-by-country basis) the measurement of progress being made in building trade capacity and the consequent trade impact in terms of improvements in the trade performance of recipient countries. To this end, two distinct series of indicators need to be considered: (i) sector-based indicators to track the sector-specific outcomes of aid-for-trade; and (ii) context indicators to measure the results of the country's policies in terms of key development goals. Building capacity is a critical component of aid-for-trade support. Proxy indicators may need to be used to measure such intangible results such as institutional building and human capacity. Since these are often country- or case-specific, the aggregation of results in trade capacity will also be a challenge.

An example: the DCED

In some aid-for-trade sectors, donors are already pursuing a harmonised approach by developing focused sets of "universal" indicators to determine levels of achievement and compare them across countries. The Donor Committee for Enterprise Development (DCED) provides an example. There is much to be learned from the DCED's experience in developing and agreeing on universal aid-for-trade indicators.

Faced with the need to show results, the members of the DCED agreed to develop common methods to quantify and measure results in private sector development (PSD)

programmes in ways that were comparable. The rationale for establishing a common standard to measure results was to:

- enable implementing organisations to quantify and communicate their achievements in ways which were credible, and which could ultimately be benchmarked;

- prevent implementing organisations from having to "reinvent the wheel", *i.e.* wasting time and energy developing a results measurement system that duplicated what others were doing and what funding organisations might later ask them to do;

- enable donors to add together and "bulk up" the results of the initiatives they fund, *e.g.* to report to their parliamentarians and tax-paying constituencies on activities to achieve the MDGs;

- support all those involved, including partner organisations, in focusing increasingly on outcomes and impacts rather than on outputs (DCED, 2010).

    - The DCED therefore agreed on and established the following three "universal" indicators, to be used in determining the level of achievements of the programme:

- *Scale:* Number of target enterprises which received financial benefits as a result of the programme's activities, each year and cumulatively. The programme was to define its "target enterprises".

- *Net income:* Additional net income (additional sales minus additional costs) accrued to targeted enterprises as a result of the programme per year. In addition, the programme was to explain why this income was likely to be sustainable.

- *Net additional jobs created:* Net additional, full-time equivalent jobs created in target enterprises as a result of the programme, per year and cumulatively. "Additional" meant jobs created minus jobs lost. The programme was to explain why these jobs were likely to be sustainable. Jobs saved or sustained could be reported separately (DCED, 2010).

There was an emphasis on limiting the number of universal indicators. Use of a few medium-term indicators would enable donors and others to aggregate and benchmark impact across different programme. It was recommended that these indicators be integrated into relevant results chains across all participating PSD programmes (DCED, 2010).

4.7 Ways forward

The need to show results in aid for trade is growing. The trade and development community is clearly shifting focus from monitoring resources to measuring outcomes. At the same time, because the Aid for Trade Initiative encompasses a broad range of activities (each with differing objectives, some not even solely trade-related), quantifying and measuring results in aid for trade is also rather complicated.

Moreover, a plethora of indicators are being generated and currently used to assess progress towards achieving specific aid-for-trade results. As more donors develop results frameworks for their aid-for-trade programmes, projects and strategies, the opportunity should be seized to move towards a more harmonised approach to monitoring and

reporting of aid-for-trade results in line with the spirit of the Paris Declaration and the Accra Agenda, which called for the maximum effort to arrive at harmonised international indicators. This requires working towards selecting a manageable number of indicators in order to avoid flooding partner countries with a sea of indicators.

This chapter has shown that there is much to gain from establishing an aligned approach to aid-for-trade results measurement. There is considerable benefit in agreeing on a limited set of common indicators across aid-for-trade programmes and projects in order to benchmark progress in building trade capacity and inform decision-making in real time, as well as to guide more in-depth analysis in evaluations. A more harmonised approach would also provide a basis for joint assessments of aid for trade with partner governments and among donors. Many different actors are involved in aid-for-trade monitoring. The process of collecting information for monitoring can therefore be made easier through a few common benchmark indicators that are central to the monitoring exercise.

This monitoring should be done at the outcome level to allow for the aggregation of results. Indeed, macro-level monitoring of aid-for-trade flows based on CRS data can be complemented by macro-level indicators of the progress being made in building the trade capacity of partner countries and the consequent trade impact in terms of improvements in these countries' trade performance. These indicators could be systematically integrated into the relevant aid-for-trade results frameworks across all donors (in addition to their regular project-specific indicators). This, in turn, would enable donors and partners alike to aggregate key aid-for-trade results for benchmarking individual countries' trade performance and for cross-country comparability.

Finally, successful application of MfDR requires a shift in mindset and focus. It needs to be based on a leadership and management culture that is focused not just on compliance, but on achieving results (including monitoring and reporting of outcomes), and that is anchored in solid evidence for decision making.

Notes

32 . *Recommendations of the Task Force on Aid for Trade* [WT/AFT/1], WTO, Geneva, 27 July 2006.

33 . "100 indicators of well-being or just one? Stiglitz v Layard." See Duncan Green's blog, "From poverty to power: how effective citizens and effective states can change the world" (*www.oxfamblogs.org/fp2p/?tag=oecd*).

34 . This section is based on *Annex 1. Promoting a harmonised approach to management for development results: Core principles*, Second Roundtable on Management for Results, Marrakech, Morocco, 2004 (*www.mfdr.org/documents/2CorePrinciples05feb04.pdf*).

35 . Based on an OECD/DAC draft paper, "Results based management in development co-operation agencies: A review of experience, Executive summary," presented at a meeting of the DAC Working Party on Aid Evaluation in February 2000.

36 . A performance-based budget implies allocating the available resources to activities that will contribute to the achievement of the desired results. It may also involve the re-allocation of resources according to "actual" results, shifting resources towards better-performing activities and away from poorer-performing ones.

37 . Chapter 1 provides a more nuanced discussion on different approaches, methods and processes for evaluating aid for trade, including conceptual challenges, particularly the linkages in results chains, the attribution problem, and the sequencing of different programmes and projects.

38 . These CoPs are: the African Community of Practice, AfCoP-MfDR (*http://copmfdrafrica.ning.com*); the Asia-Pacific Community of Practice, AsCoP-MfDR (*http://cop-mfdr.adb.org*); and the Latin American and Caribbean Community of Practice, CoPLAC-MfDR (*www.iadb.org/PRODEV/CoPLAC-MfDR.cfm*).

39 . "Managing for Development Results, MfDR Capacity Scan" (*www.mfdr.org/CAP-Scan.html*) .

40. Donor agencies generally use both quantitative and qualitative indicators. Quantitative indicators have the potential to calibrate the results of activities in terms of volume and, on that basis, facilitate the adoption of a cost-effective approach. Qualitative indicators can be used to assess behaviour and attitudes with a view to measuring the impact and sustainability of the results.

41 . See, for instance, Annex 2, Table 2, in OECD (2007), *Trade-Related Assistance: What Do Recent Evaluations Tell Us?*, which provides an overview of performance indicators used by various donor agencies to assess different types of trade-related assistance (*www.keepeek.com/Digital-Asset-Management/oecd/development/trade-related-assistance_9789264031203-en*).

42. The "soundness" of a results-based monitoring framework is judged based on three criteria: (i) the quality of the information generated; (ii) stakeholder access to the information; and (iii) the extent to which the information is utilised within the country.

43. The study, "Results Reporting by Donor Agencies, Draft Final Report (1 June 2010)", was undertaken by consultants (MDF and Goss Gilroy Inc.) for the OECD-DAC Global Partnership for MfDR.

44 . "EuropeAid Evaluation Guidelines: Evaluation Methods" (*http://ec.europa.eu/europeaid/evaluation/methodology/methods/mth_cid_en.htm*).

45 . The DAC defines indicators as a "quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor" (OECD, 2002). They need to be simple, but also SMART (specific, measurable, attainable, relevant and time-bound) and comparable across countries (OECD, 2008b). Moreover, indicators should be used to "provide approximate answers to a few important questions rather than seek to provide exact answers to many less important questions" (UNDP, undated).

46 . UNDP presentation on "The Case for Indicators of Context, Trade Mainstreaming and Donors' Response" at the WTO Symposium on Identifying Indicators for Monitoring Aid for Trade, 15-16 September 2008 (*www.wto.org/english/tratop_e/devel_e/a4t_e/symp_sept08_sess3_undp_e.pps*).

47 . A few donors, such as the UK Department for International Development (DFID), the Inter-American Development Bank (IDB) and the World Bank, have started to develop results frameworks for their aid-for-trade programmes and strategies.

48 . Trade capacity building covered the trade policy environment, policy-making capacities relevant to national, regional and multilateral trade, export-related capacities and infrastructure, trade facilitation and support services, and market access (OECD, 2001).

49 . See: *www.oecd.org/document/5/0,3343,en_2649_34665_39119685_1_1_1_1,00.html*

References

Binnendijk, A. (2000), *Results Based Management in the Development Co-operation Agencies: A Review of Experience, Executive Summary*, a draft report prepared for the meeting of the OECD DAC Working Party on Aid Evaluation, February 2000.

Brenton, P. and E. von Uexküll (2009), "Product specific technical assistance for exports – has it been effective?", *Journal of International Trade and Economic Development*, Vol. 18, No. 2.

Calì, M. and D.W. te Velde (2009), "Does aid for trade really improve trade performance?", http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1430492.

DCED (Donor Community for Enterprise Development) (2010), *Measuring Achievements in Private Sector Development: Implementation Guidelines*.

Elliot, K.A. (2007), "Ensuring that Aid for Trade promotes development: Indicators for measuring the impact", unpublished.

European Commission (2006), *Evaluation Methods for the European Union's External Assistance, Methodological Bases for Evaluation, Volume 1*, Joint Evaluation Unit, EuropeAid Co-operation Office, DG Development, DG External Relations, Brussels.

Gamberoni, E. and R. Newfarmer (2009), "Aid for Trade: Matching Potential Demand and Supply", *World Bank Policy Research Working Paper 4991*.

GTZ (German Technical Cooperation Agency) (2008), *Results-based Monitoring: Guidelines for Technical Cooperation*, Corporate Development Unit, Deutsche Gesellschaft für Technische Zusammenarbeit GmbH, Eschborn.

Hoekman, B. and A. Nicita (2010), "Assessing the Doha Round: Market access, transaction costs and aid for trade facilitation", *Journal of International Trade and Economic Development*, Vol. 19, No. 1.

JICA (Japan International Co-operation Agency) (2005), *Handbook for Selecting Outcome Indicators: A Guide to Practical Evaluation of Technical Co-operation*.

Manning, R. (2009), *Using indicators to encourage development: Lessons from the Millennium Development Goals*, DIIS Report 2009:01, Danish Institute for International Studies, Copenhagen

MCC (Millennium Challenge Corporation) (2009), *Policy for Monitoring and Evaluation of Compacts and Threshold Programs*.

OECD (2001), *The DAC Guidelines: Strengthening Trade Capacity for Development*. OECD, Paris.

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

OECD (2007), *Trade-Related Assistance: What Do Recent Evaluations Tell Us*? OECD, Paris. ww.oecd.org/dataoecd/31/45/37624158.pdf

OECD (2008a), *2008 Survey on Monitoring the Paris Declaration: Making Aid More Effective by 2010*. OECD, Paris.

OECD (2008b), *Evaluating Development Co-operation: Summary of Key Norms and Standards*. OECD, Paris.

OECD (2009), *Managing for Results in Aid for Trade*, OECD, Paris, *http://www.oecd.org/officialdocuments/displaydocumentpdf/?cote=COM/DCD/TAD(2009)4&doclanguage=en*.

OECD (2010a), *How to Evaluate Aid for Trade: Approaches, Methodologies and Processes*, COM/DCD/TAD(2010)2. OECD, Paris.

OECD (2010b), "Trade Facilitation Indicators: the Impact on Trade Costs", TAD/TC/WP(2010)5. OECD, Paris.

OECD/World Bank (2006), *Sourcebook: Emerging Good Practices in Managing for Development Results*, first edition. www.mfdr.org/Sourcebook/1stEdition/MfDRSourcebook-Feb-16-2006.pdf

OECD/WTO (World Trade Organization) (2009), *Aid for Trade at a Glance 2009: Maintaining the Momentum*. OECD, Paris.

Smith, M.G. (2009), *Monitoring and Evaluation in the Enhanced Integrated Framework: Draft Policy and Guidelines*, Final Draft, 12 February.

Toffolon-Weiss, *et al.* (1999), "The Results Framework – An innovative tool for program planning and evaluation", Evaluation Review, Vol. 23, No. 3.

Wiebe, F.S. (2008), *Aid Effectiveness: Putting Results at the Forefront, MCC's "New Institutional Approach"*, Millennium Challenge Corporation, Washington, D.C.

WTO (2006), *Recommendations of the Task Force on Aid for Trade*, WT/AFT/1, 27 July 2006, Geneva.

Annex A.
The OECD meta-evaluation: overview of evaluations

Overview of the evaluations

Table A1 (A) provides a broad overview of the evaluations analysed (number of evaluations, number of pages, average number of pages). While it is possible to compare the data on Ghana and Vietnam, the transport and storage sectors have been treated as one sector (*i.e.* Transtor). A distinction has been made between the overall (162) and narrow (42) sets of evaluations. In both the overall and narrow sets, the number of evaluations and the number of pages in the evaluations suggest that programmes and projects in Vietnam were evaluated more intensively than those in Ghana. However, the average number of pages ("average length of evaluations") in the overall and narrow sets was greater in the case of Ghana.

Table A1. Overview of evaluations (1999-2009)[1]

	Ghana			Vietnam			TranStor	
	overall	narrow	Period	overall	narrow	Period	overall	narrow
A. Data on evaluations								
Number of evaluations	34	13	1999-2009	64	9	1999-2009	64	20
Number of pages	3,312	1,858	1999-2009	5,719	814	1999-2009	4,227	1,438
Average length of evaluations	97	143	1999-2009	89	90	1999-2009	66	72
B. Trade data								
Exports (current USD billion)	5.3		2008	62.7		2008	--	
Imports (current USD billion)	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	--	
C. Macroeconomic data								
Population (million)	23.3		2008	86.2		2008	--	
GDP (current USD billion)	16.1		2008	90.7		2008	--	
GDP (at PPP USD billion)	33.9		2008	240.1		2008	--	
GDP/capita (PPP USD)	1,454.9		2008	2,785.3		2008	--	
Real GDP growth index	153.0		(2000=100)	179.0		(2000=100)	--	
Gross external debt (USD billion)	5.8		2008e	23.7		2008e	--	
Public debt (% GDP)	67.5		2008e	44.5		2008e	--	
Economic aid (USD billion)	1.65		2006-2009	2.95		2006-2009	--	
D. Poverty data								
Below poverty (% population)	28.5		2007	7.8		2009e	--	
Gini index	39.4		2005	37.0		2004	--	

Transtor: transport and storage sectors combined

PPP: purchasing power parity; 2008e: estimated figure for 2008; 2009e: estimated figure for 2009

Source: DEReC, WTO. Authors' computations.

These observations need to be put into perspectives by taking into account basic differences between the Vietnamese and Ghanaian economies. They can be defined in terms of:

- relative size of trade flows (Table A1, B). This definition echoes the pure "trade" aspect of aid for trade. While Ghana's exports and imports are much smaller (between one-fourth and one-seventh of Vietnam's), its programmes and projects appear to be more intensively monitored.

- relative size of populations, GDP and other key macroeconomic variables (*e.g.* debt and economic aid) (Table A1, C). Each of these variables shifts the focus from the pure "trade" aspect of aid for trade to indicators more directly related to development and aid. Again, despite having smaller macroeconomic indicators (roughly one-third to one-fourth of those of Vietnam), trade-related programmes and projects in Ghana are more intensively monitored compared to Vietnam – on average two to three times more intensely.

- differences in income distribution and the share of the population below the poverty level or the Gini coefficient (Table A1, D). As a higher percentage of the population of Ghana is below the poverty level, Vietnam appears to be subject to more evaluations relative to Ghana. This result raises interesting questions, always assuming the sample of evaluations is representative. Does it reflect the different nature of the programmes and projects being evaluated or of the wider aid policy? Or does it reflect the countries' specificities? For instance, is it easier to undertake programmes and projects in Vietnam than in Ghana? Or does it reflect evaluators' greater sensitivity to key development concerns (*e.g.* income distribution) if a country has reached a certain level of development and seems to be on a successful growth path?

Different types of programmes, projects and general aid policies

The nature of the programmes and projects evaluated varies considerably, making them more or less easy to evaluate. In Table A2 the overall and narrow sets of evaluations are broken down into three main categories. Following the OECD/DAC definitions, 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*" while projects involve "*well-defined development activities that are limited in time and space*". For the purposes of this publication, the complete body of operations by a specific donor in a specific recipient country within a particular time frame may be referred to as "general aid policy".

The different scopes of programmes, projects and general aid policy help to determine the characteristics of evaluations. Evaluations of projects appear less easy to relate to broad trade and development issues than those of programmes and general policy operations for a number of reasons. Projects are also likely to be more limited in terms of their time horizons (*i.e.* to be shorter-term) and there may not be sufficient time to assess their consequences. On the other hand, evaluating the achievements of projects may be methodologically easier than evaluating those of a general aid policy, with much more ambitious and longer-term goals.

Table A2. Breakdown of evaluations by programme, project and general aid policy (1999-2009)

	Ghana		Vietnam		Transtor	
	overall	narrow	overall	narrow	overall	narrow
A. All evaluations						
Number of evaluations	34	13	64	9	64	20
Number of pages	3,312	1,858	5,719	814	4,227	1,438
Average length of evaluations	97	143	89	90	66	72
B. Evaluations of programmes						
Number of evaluations	12	4	18	4	18	8
Number of pages	1,153	485	1,675	370	1,845	872
Average length of evaluations	96	121	93	93	103	109
C. Evaluations of projects						
Number of evaluations	11	2	27	5	38	12
Number of pages	670	114	2,143	483	1,949	566
Average length of evaluations	61	57	79	97	51	47
D. Evaluations of general aid policy operations						
Number of evaluations	11	7	19	--	8	--
Number of pages	1,489	1,259	1,901	--	504	--
Average length of evaluations	135	180	100	--	63	--

Transtor: transport and storage sectors combined

Source: DEReC. Authors' computations.

Time patterns of evaluations, by recipient

This section focuses on the time pattern of evaluations according to recipient. It looks first at the overall set of evaluations, and then at the narrow set.

Overall set

Table A3 shows that the time pattern for the overall set of evaluations varies widely across the countries and sectors considered. There is greater variation in the number of evaluations, number of pages and average length of evaluations (*i.e.* number of pages) in the case of Ghana than in that of Vietnam, although the number of evaluations for both countries peaked in the period 2004-2006. This volatility can be measured using an indicator of variance. The indicator chosen 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). This indicator of variance is two to four times higher in the case of Ghana than in that of Vietnam. Evaluations in the transport and storage sectors suggest a time pattern closer to the one in Ghana than to the one in Vietnam.

Nevertheless, it is essential to go beyond the number of evaluations or number of pages, and to provide a measure which reflects these evaluations' potential informational content. As explained in Chapters 2 and 3, the frequency with which the 48 key words and expressions are used (*i.e.* their occurrence per 100 pages) provides an – admittedly crude – measure of the evaluations' potential informational content.

Table A3 shows the frequency by recipient for the overall set of 48 key words and expressions. The average frequency is notably lower for Ghana than for Vietnam or for the transport and storage sectors. The variance is greater for Ghana than for Vietnam or for the transport and storage sectors.

Table A3. Overview of the overall set of evaluations: time pattern by recipient (1999-2009)

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	All years	Weight st. dev.
Ghana													
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	3,312	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
Vietnam													
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	5,719	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
Transtor													
Number of evaluations	2	11	3	11	1	12	6	7	7	4	0	64	0.72
Number of pages	100	1,044	112	549	100	498	302	466	812	244	--	4,227	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

st. dev. = standard deviation

Transtor: transport and storage sectors combined

Note: [a] Frequency with which key words and expressions were used. See Table 2.1 for the list of these words and expressions.

Source: DEReC. Authors' computations.

These observations raise the following question: Are the reasons for such differences related to the recipient countries or to the donors? Some of the reasons may be related to the recipient.

In the case of Ghana, "at the border" policies were subject to two forces during this decade (WTO Trade Policy Review, 2001 and 2008). The simple average applied tariff fell from 17% in 1992 to 14.7% in 2000 and 12.7% in 2007. However, Ghana's tariffs are still largely unbound in the WTO, with a very low binding coverage of 14.3% (Table A1), covering comprehensively (but practically only) agricultural products. This means that at any time tariffs could be significantly raised, with little risk of retaliation by trade partners. For instance, in February 2000 Ghana introduced a "special import tax" of 40% on a significant share of consumer goods (WTO Trade Policy Review, 2001).

Moreover, since 2009 the world consensus on open trade policies has come under stress because of the global financial crisis and recession. Because of Ghana's low binding coverage, it retains the option of increasing its applied tariffs at any time without risk of breaking its WTO commitments or being forced to provide compensation to its trading partners. A 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).

The trajectory of Vietnam's "at the border" and "behind the border" policies is much more ambitiously liberal (partly due to having been largely a closed economy in the

1980s) and appears firmer than Ghana's. In trade matters, the decade under consideration was dominated by Vietnam's WTO accession negotiations. Its Protocol of Accession was ratified in 2007 following 12 years of negotiations (a relatively short period). The Protocol includes commitments on a wide range of "at the border" and "behind the border" policies. Vietnam has agreed to bind all its tariffs at an average rate of 11.4% (Table A1) and to ambitious discipline on non-tariff barriers (from quotas to standards), foreign direct investment, services and intellectual property. This is in sharp contrast to Ghana, which in many policy areas and sectors has made no commitment. 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

The overall set of evaluations is heterogeneous, with many evaluations only loosely related to aid for trade. Table A4 is an overview of the narrow set of 42 evaluations.

Table A4. Overview of the narrow set of evaluations: time pattern by recipient (1999-2009)

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	All years	Weight st. dev.
Ghana													
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	1,858	0.70
Avg length of evaluations	528	134	--	--	--	--	68	140	--	104	58	143	1.03
Frequency [a]	546	801	--	--	--	--	1,061	736	--	1,513	336	763	0.50
Vietnam													
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	--	--	1,274	629	1369	828	1134	0.40
Transtor													
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	--	1,438	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	1,000	1019	--	819	0.32

St. dev. = standard deviation

Transtor = transport and storage sectors combined

Note: [a] Frequency with which key words and expressions were used. See Chapter 2, Table 2.1 for a list of these words and expressions.

Source: DEReC. Authors' computations.

Evaluations by donor

In this section the focus switches from recipients to donors. The evaluations by the Dutch Ministry of Foreign Affairs' Policy Operation and Evaluation Department (IOB), the European Bank for Reconstruction and Development (EBRD), NEI and the Norwegian Agency for Development Cooperation (NORAD) have been excluded from the sets of evaluations because these four donors reported to DEReC only one evaluation for the three fields (Ghana, Vietnam and Transtor). Once again, a distinction is made between the overall and narrow sets of evaluations.

Overall set of evaluations

Several conclusions may be drawn from Table A5. 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".[50] From this perspective, there is a marked difference between Vietnam and transport and storage, on one hand, and Ghana on the other. Evaluations of Vietnam and transport and storage are largely dominated by one or two donors: SIDA and, to a lesser extent, ADB for Vietnam (75.9% of total pages); ADB and (to a notably smaller extent) AfDB and the World Bank (76.5% of total pages) for transport and storage. By contrast, the Ghana evaluations reveal no dominant donor. Four donors (AfDB, DANIDA, DFID and the World Bank) have almost the same share of pages (15-17%) although AfDB remains important with regard to the number of evaluations.

This difference between a supply structure dominated by a couple of donors and one fragmented among several is an interesting feature – one that might warrant further investigation as to whether these differing supply structures contributed to the large dispersion in evaluations for Ghana compared to those for Vietnam. However, 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.

Second, Table A5 gives a sense of the evaluations' potential informational content by donor through 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, and six for Vietnam) among the donors exhibiting the highest and lowest frequencies. The relative ranking of all the donors shown (*e.g.* those with frequencies higher than the average frequency for the overall set of evaluations) is relatively stable. The same observation can be made if one limits the donors presented to those with the highest number of evaluations. This stability of results 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 for transport and storage, with a factor of only 1.5.

These results are based on the overall set of 48 key words or expressions. The complete list of words and expressions is presented in Chapter 2, Table 2.1. It consists of three groups (roughly similar in size) of words and expressions related to trade, development, and procedures and techniques.

Table A5. Overview of the overall set of evaluations: donor pattern

	AfDB	ADB	AFD	CIDA	DANIDA	DFID	EC	GTZ	Japan	SECO	SIDA	USAID	WB	J-Eval	ALL	Weighted st. dev.
Ghana																
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	1,132	900	779	1,167	--	812	--	513	487	715	--	868	0.35
core list	397	--	72	587	362	382	373	--	419	--	191	149	384	--	336	0.46
core 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
Vietnam																
Number of evaluations	--	19	2	1	--	1	2	2	1	3	28	--	2	3	64	1.55
Number of pages																
total	--	1,790	80	34	--	83	162	120	13	130	2,721	--	153	433	5,719	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	1,035	--	906	1,515	600	708	987	484	--	827	961	687	0.37
core list	--	340	44	538	--	439	386	273	277	268	161	--	348	425	257	0.43
core 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
Transtor																
Number of evaluations	9	35	--	--	3	--	--	1	1	--	3	--	9	--	61	1.39
Number of pages																
total	485	2179	--	--	203	--	--	6	19	--	290	--	555	--	3,737	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
core list	416	293	--	--	293	--	--	283	126	--	169	--	367	--	295	0.37
core loss rate	53.2	45.6	--	--	45.6	--	--	33.3	17.6	--	33.5	--	45.8	--	43.8	0.30

WB = World Bank; J-Eval = joint evaluation; st. dev. = standard deviation

Note: [a] Frequency with which key words and expressions were used. See Chapter 2, Table 2.1for a list of these words and expressions.

Source: DEReC. Authors' computations.

It seems useful to refine this list by focusing on subsets of words and expressions considered particularly important ("core"). Such a core list could be defined as the 28 words and expressions considered essential from a purely economic perspective (Table A6). In other words, it would give a sense of the evaluations which are most clearly driven by economic analysis.

Table A6. Core list of words and expressions

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

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

The narrow set of evaluations

The narrow set relies on a substantially smaller number of observations (including in terms of donors). This means the observations made should 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 A7. Overview of the narrow set of evaluations: donor pattern

	AfDB	ADB	AFD	CIDA	DANIDA	DFID	EC	GTZ	Japan	SECO	SIDA	USAID	WB	J-Eval	All donors	Weighted st. dev.
Ghana																
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
Vietnam																
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]	--	1,260	--	--	--	--	1,783	580	--	987	736	--	--	--	1134	0.44
Transtor																
Number of evaluations	12	3	--	--	1	--	--	--	1	--	--	--	3	--	20	1.15
Number of pages																
total	942	160	--	--	47	--	--	--	19	--	--	--	270	--	1,438	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

WB = World Bank; J-Eval = joint evaluation; st. dev. = standard deviation

Note: [a] Frequency with which key words and expressions were used. See Chapter 2, Table 2.1 for a list of these words and expressions.

Source: DEReC. Authors' computations.

Notes

[50] The number of pages seems a more accurate indicator than the number of evaluations.

Annex B.
Examples of existing trade-related indicators

Trade Performance Indicators		Source
Trade and Development Index (TDI)	The TDI is designed as a mechanism for monitoring the trade and development performance of countries; a diagnostic device to identify factors affecting such performance; and a policy tool to help stimulate and promote national and international policies and actions for development and poverty reduction.	United Nations Conference on Trade and Development (UNCTAD)
Trade Performance Index (TPI)	The TPI calculates the level of competitiveness and diversification of a particular export sector and compares results across countries. Currently the TPI covers 184 countries and 14 different export sectors. It provides three types of indicators: a general profile; a country position for the latest available year; and changes in export performance in recent years. Its composite ranking is based on five criteria, which are value of net exports, per capita exports, world market share, and diversification of products and of markets.	International Trade Centre
Enabling Trade Index (ETI)	The ETI is an aggregate indicator constructed from a range of both hard data and survey data. It focuses on the broader trading environment in a country. The ETI aims to assess the extent to which countries have in place the institutions and policies for enabling trade. The World Economic Forum publishes an annual report in which 123 different countries are measured against this index.	The World Economic Forum
World Trade Indicators (WTI)	The WTI database is a tool that enables countries to benchmark their trade policy and performance and compare across countries and country groupings (*e.g.* by region, income group, regional trade agreements). It contains a broad set (about 450 variables) of trade-related policy and outcome indicators for 211 countries and territories. WTI is organised around five thematic pillars: trade policy; external environment; institutional environment; trade facilitation; and trade outcome.	World Bank Institute
Category-specific Indicators		
Infrastructure and Agribusiness Indicators	In 2005 the IFC established a systematic indicator framework called the Development Outcome Tracking System (DOTS) to monitor the development results of its investments and advisory services and make performance comparisons across its projects. Within this results tracking system, each IFC industry department has identified a number of standard, industry-specific indicators that are tailored to focus on those outcomes that are most relevant to each industry.	IFC (International Finance Corporation)
AICD database on Africa's Infrastructure	The AICD programme has developed a suite of indicators (containing a total of 893 variables) to measure performance in nine major infrastructure sectors (air transport, ICT, irrigation, ports, power, railways, roads, water and sanitation) across 24 African countries. Quantitative indicators include infrastructure performance measures of access, efficiency, quality and financial performance, with a focus on infrastructure service providers such as utilities. Qualitative indicators measure the institutional, legal and regulatory frameworks of each sector.	AICD (Africa Infrastructure Country Diagnostic)
Logistics Performance Index (LPI)	The LPI is a benchmarking tool focused specifically on measuring the trade and transport facilitation "friendliness" of countries. It reflects the overall perception of a country's logistics environment and compares the trade logistics profiles of 155 countries. LPI measures: efficiency of the customs clearance process; quality of transport and transport-related infrastructure; ease of arranging competitively priced shipments; competence and quality of logistics services; ability to track and trace consignments; and frequency with which shipments reach the consignee within the scheduled or expected time.	World Bank
Doing Business; Trading Across Borders Indicators	The Trading Across Borders indicator series (drawn from the Doing Business database) represents a country's trade facilitation capabilities and consists of objective measures of the trade facilitation environment: number of documents for import and export; time (in days) for import and export; and cost (USD per container) to import and export. It estimates the monetary costs associated with shipping goods from the factory gate to ports, and from ports to retail outlets for a standard container.	World Bank
World Telecommunication/ICT Indicators Database	This database captures the level of advancement of information and communications technologies in more than 150 countries. Its main objective is to provide policy makers with a tool to benchmark and assess their information society developments and to monitor progress	International Telecommunication Union

	made globally to close the digital divide.	
Travel and Tourism Competitiveness Index (TTCI)	The TTCI, which covers 133 countries, provides a comprehensive strategic tool for measuring the factors and policies that make it attractive to develop the tourism sector in different countries. The index is based on over 70 indicators.	World Economic Forum
Indicators of Financial Structure, Development and Soundness	These indicators include system-wide indicators of size, breadth and composition of the financial system; indicators of key attributes such as competition, concentration, efficiency and access; and measures of the scope, coverage and outreach of financial services.	IMF (International Monetary Fund)
IEA Energy Statistics	The Energy Statistics Division of IEA collects, processes and publishes data and information on energy production, trade, stock, transformation, consumption, prices and taxes as well as on greenhouse gas emissions. Data are available for all OECD member economies and over 100 non-OECD member economies.	IEA (International Energy Agency)
Asia Regional Integration Indicators	The Asia Regional Integration Centre has developed a set of indicators which measure regional integration in areas of trade, investment and financial markets. Data are collected from 19 Asian countries.	ADB (Asian Development Bank)

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new development and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where government can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia; Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg Mexico, the Netherlands, New Zealand, Norway, Poland, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Commission takes part in the work of the OECD.

OECD publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.