BMZ | Federal Ministry for Economic Cooperation and Development

50 Years
Building the future.
Let's join forces.

Tools and Methods for Evaluating the Efficiency of Development Interventions

Foreword by the Federal Ministry for Economic Cooperation and Development

BMZ Evaluation Working Papers address a broad range of methodological and topical issues related to the evaluation of development cooperation. Some papers complement BMZ evaluation reports; others are free-standing. They are meant to stimulate discussion or serve as further reference to published reports.

Previous BMZ Evaluation Working Papers have focused on measuring impact. The present paper explores approaches for assessing efficiency. Efficiency is a powerful concept for decision making and ex post assessments of development interventions but, nevertheless, often treated rather superficially in project appraisal, project completion and evaluation reports. Assessing efficiency is not an easy task but with potential for improvements, as the report shows. Starting with definitions and the theoretical foundations the author proposes a three level classification related to the analytical power of efficiency analysis methods. Based on an extensive literature review and a broad range of interviews, the report identifies and describes 15 distinct methods and explains how they can be used to assess efficiency. It concludes with an overall assessment of the methods described and with recommendations for their application and further development.

This paper has been produced by Markus A. Palenberg, the managing director of the Institute for Development Strategy (IfDS), Munich, and a fellow at the Global Public Policy Institute in Berlin. The opinions presented in this study are those of the author(s) and do not necessarily reflect the views of the BMZ.

This paper is available online at
http://www.bmz.de/en/what_we_do/approaches/evaluation/Evaluation/methods/index.html.
It should be cited as follows: Palenberg, M. (2011): Tools and Methods for Evaluating the Effi-ciency of Development Interventions. *Evaluation Working Papers.* Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung.

Michaela Zintl
Head of Division "Evaluation of Development Cooperation; Auditing"
Federal Ministry for Economic Cooperation and Development

Acknowledgements

---

[1] The first person plural pronoun „we" is used throughout this report in order to acknowledge the substantial participation of interviewees (especially of those listed in this acknowledgment), to avoid passive language and out of habit. Nevertheless, all observations marked with „we" are the strictly the author's own and opinions of others are labelled as such.

Contents

Annexes

Abbreviations and Acronyms

Institutions

AFD	Agence Française de Développement (French Development Agency)
ANR	Agence Nationale de la Recherche (French National Research Agency)
BMZ	German Federal Ministry for Economic Cooperation and Development
ECDPM	European Centre for Development Policy Management
EuropeAid	EuropeAid Co-operation Office of the European Commission
GTZ[2]	Deutsche Gesellschaft für Technische Zusammenarbeit
IEG	World Bank Independent Evaluation Group
IFAD	International Fund for Agricultural Development
InWEnt[2]	Capacity Building International
KfW	KfW Entwicklungsbank (part of the KfW Bankengruppe)
MCC	Millennium Challenge Corporation
OECD DAC	OECD Development Assistance Committee (OECD: Organisation for Economic Co-operation and Development)
QAG	World Bank Quality Assurance Group
SADEV	Swedish Agency for Development Evaluation
UNIDO	United Nations Industrial Development Organization
UTV	Sida's department for evaluation (Sida: the Swedish International Development Cooperation Agency)
WB	The World Bank
WBI	World Bank Institute

Methods and Technical Terms

CBA	Cost-Benefit Analysis
CEA	Cost-Effectiveness Analysis
ceteris paribus	holding other factors constant
CUA	Cost-Utility Analysis
DALY	Disease-Adjusted Life Year
DFE	Discounted Future Earnings. Used as "DFE method" in this report.

---

[2] DED, GTZ and InWEnt were merged into Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) on 1 January 2011.

ERR	Economic Rate of Return
GDP	Gross Domestic Product
GRPP	Global and Regional Partnership Program
IRR	Internal Rate of Return
MADM	Multi-Attribute Decision-Making
ME	Effects Method (Méthode des Effets)
NPV	Net Present Value
QALY	Quality-Adjusted Life Year
RC	Required Compensation. Used as "RC approach" in this report
ROI	Return on Investment
SWAp	Sector-Wide Approach

Executive Summary

This report investigates tools and methods for assessing aid efficiency. It explains concepts of efficiency and efficiency analysis and presents a catalogue of 15 methods that can be used to assess the efficiency of aid interventions. Each method is described and characterised. Several general observations and recommendations conclude the report.

Motivation. The overall study was motivated by the apparent gap in evaluation studies between expected and delivered results in terms of efficiency analysis of aid interventions.

- On the one hand, efficiency assessment is a principal aid evaluation criterion, required by bilateral and multilateral policies;

- On the other hand, several reports and observations reveal that donors and development agencies conduct efficiency analyses with insufficient frequency and quality.

This report's catalogue of methods for conducting efficiency analyses provides assistance in closing this gap.

Definitions of efficiency. The OECD Development Assistance Committee defines efficiency in terms of *transformation of inputs into results*. Similarly, welfare economists sometimes define efficiency based on the *transformation of costs into benefits* as measured, for example, by benefit-cost ratios. In both cases, *transformation efficiency* is measured by how economically inputs or costs are transformed into results or benefits.

In *welfare economics*, costs and benefits are understood in general terms and include social and private, direct and indirect, and tangible and intangible contributions.

If all inputs and results are taken into account, transformation efficiency is sometimes referred to as *allocation efficiency*. If only results on the output-level are taken into account, the related transformation efficiency measure is called *production efficiency*.

Transformation efficiency is often measured by ratios. In the case of allocation efficiency, benefit-cost ratios, cost-effectiveness ratios, cost-utility ratios and internal and economic rates of return are used. In the case of production efficiency, unit costs or other partial efficiency indicators are used.

Apart from accounts of *transformation efficiency*, welfare economists also describe efficiency based on optimisation principles. For example, *Pareto improvements* increase the individual welfare of some people without making others worse off. *Kaldor-Hicks improvements* extend this principle and allow for special types of trade-offs, i.e. the compensation of welfare losses of one person by welfare gains of another person.

The optimisation rule most frequently applied in welfare economics is based on the concept of *net benefits*, i.e. the difference between benefits and costs. Net benefits describe the overall impact an intervention has on welfare, and optimisation efficiency would then be measured by net benefits.

In *decision analysis* – a different field of academic and applied research – the concept of net benefits is generalised to the concept of *utility* that measures a decision-maker's relative preferences with regard to different options and their consequences.

In both cases, efficiency is measured by the degree to which optimisation rules are fulfilled. We refer to efficiency defined in this way as *optimisation efficiency*.

The distinction between optimisation and transformation efficiency appears superfluous at first: aren't these simply two different measures for the same thing? Sometimes, they are, but often, they are not. For example, choosing the most efficient development interventions based on transformation efficiency information may lead to different results than selecting interventions based on their optimisation efficiency.

In this report, we consider efficiency assessment methods based on transformation efficiency as well as on optimisation efficiency.

We consider methods that provide only partial accounts of efficiency, in the sense that they consider only a subset of inputs and results or costs and benefits. While many methods require quantitative data, we also consider approaches that deal with efficiency in entirely qualitative ways.

Quality of efficiency analysis. The measurement of efficiency may require special skills and experience. Carol Weiss, a well-known evaluation expert, describes efficiency analysis as *a specialized craft that few evaluators have mastered*. All methods presented in this report face principal and practical challenges and limitations that must be considered when applying them.

Principal challenges and limitations depend on the approaches used. For example, for methods based on welfare changes to society, results depend on the specific social welfare function chosen or on the model used for estimating and aggregating individual welfare.

Practical challenges and limitations arise from the approximations made by evaluators within specific methods, as well as from the quality and availability of data.

In addition, basic requirements for sound evaluation are required. These include, for example, the measurement of net effects

against a reference scenario. Otherwise, effects caused by an intervention and other changes that have not been caused by an intervention are mixed up.

Efficiency as rationale for decision-making. Efficiency is a powerful concept. At least in theory, welfare can be maximised based on efficiency information alone and efficiency would therefore represent the most important criterion in appraisals and evaluations.

In practice, however, challenges such as the limited scope of efficiency analysis, simplifications and approximations reduce its potential.

Therefore, efficiency analysis usually does not dictate decisions but can provide crucial information for decision-makers. Even without accurate efficiency-related information, the concept of efficiency remains important for informing a welfare-maximising approach to aid.

Analysis perspectives. As with other evaluation methods, the result of efficiency analyses depends on the *analysis perspective*. Important analysis perspectives are those of the entire aid recipient society or of a part of that society, for example the beneficiaries of an aid intervention.

However, other, more restricted analysis perspectives can be useful as well:

- Decision-making analysis is usually based on the perspective of a single person, the decision-maker. He or she takes into account other analysis results and the opinions and preferences of others, but nevertheless evaluates options from his or her individual perspective.

- Analysis perspectives of private sector entities, e.g. companies, are useful for determining the viability of a business element embedded in an aid intervention.

Probably the most comprehensive analysis perspective covers not only the aid recipient society but also elements of the society providing aid, for example by explicitly considering the donor agency as part of the analysis. In this way, the operational efficiency of that agency is included in the assessment of the aid interventions' efficiency.

Without explicitly or implicitly describing analysis perspectives, results of efficiency analyses are not only difficult to interpret, but can also not be compared with each other and lead to confusion.

Criteria for characterising efficiency analysis methods. In order to describe and assess efficiency analysis methods, we consider their *analytic power* as well as the *analysis requirements* in terms of data, time and skills.

In our definition, analytic power is essentially determined by the *analysis level* of a method. We differentiate between three levels of analysis:

- *Level 2 analysis*, the most potent, is capable of assessing the efficiency of an aid intervention so that it can be compared with alternatives or benchmarks.

- *Level 1 analysis* is capable of identifying the potential for efficiency improvements within aid interventions. Level 1 and 2 analyses have complementary functions: while level 2 analyses support the selection of interventions, level 1 analyses primarily help to improve interventions operationally.

- Finally, *level 0 analysis* is entirely descriptive and can usually not produce well-founded recommendations. For this level of analysis, recognising the limitations of the findings is critical to avoid proceeding with actions under a misconception of the evidence basis.

The assessment of analytic power is complemented by the degree to which methods are well-defined, the degree to which different evaluators can be expected to produce the same analysis results (if all other things remain equal), and the way stakeholders are involved in the analysis.

In addition, methods are characterised by their data, time and skill requirements:

- Data requirements are assessed both by the type and the origin of data.

- Basic data types are *qualitative* information and *quantitative* data. The latter is further subdivided into *financial* and non-financial (*numerical*) data. Some methods express costs and benefits originally measured in any of these three data types in monetary units. We use the additional qualifier *monetarisable data* to refer to this case.

- With data origin, we describe what level of an intervention's results chain the data stems from, i.e. input, output or outcome and impact level data.

- Time requirements for conducting efficiency analyses are measured in terms of working times for both the evaluator and stakeholders.

- Finally, skill requirements indicate whether skills needed for the analysis exceed what we consider basic evaluation skills

Overview of methods. Overall, 15 distinct analysis methods have been identified and are described in this report. The following table provides an overview of these methods, ordered according to the analysis level and the degree to which methods were known by the experts interviewed for this study.

.

Degree to which method is known	Level 2 methods	Level 1 methods	Descriptive methods
Well-known methods	Cost-Benefit Analysis (CBA)	Benchmarking of unit costs	Expert judgement
Somewhat less well-known methods	Cost-Effectiveness Analysis (CEA)	Follow the Money	
		Financial analysis	
		Stakeholder-driven approaches	
		Benchmarking of partial efficiency indicators other than unit costs	
Methods unknown to a substantial fraction of evaluation experts	Multi-Attribute Decision-Making (MADM): Intuitive scoring models	Comparative ratings by stakeholders: Comparative rating of effectiveness and cost analysis	Specific evaluation questions on efficiency
	Multi-Attribute Decision-Making (MADM): Scientific decision analysis		
	Effects Method	Comparative ratings by stakeholders: Comparative rating of efficiency	
	Cost-Utility Analysis (CUA)		

Conclusions. Based on the findings of this report, we draw four general conclusions. The first two illustrate how efficiency analyses can be applied more widely:

- First, the application potential of efficiency analysis methodology is not exhausted, both in terms of frequency and quality.

  Regarding the frequency of application, we have identified several methods that are little known and sometimes not well-documented. They are applicable in circumstances where more established methods are not suitable. In other circumstances, these methods can complement better-known methods. Examples are Cost-Utility Analyses, Methods for Multiple-Attribute Decision-Making, and more pragmatic methods such as comparative ratings by stakeholders and the Follow the Money approach.

  Regarding quality, various reports indicate that efficiency analysis is often applied with insufficient rigour.

To successfully conduct an efficiency analysis, advanced analytic skills may be required, both for the evaluator conducting the analysis and for those commissioning an evaluation or an appraisal. Without appropriate skills, impractical or inappropriate methodology for efficiency analysis may be selected and guidance and quality control may be weak.

In addition, for some methods, the evaluation design needs to be changed from *vertical assessments* that evaluate several criteria for a single intervention to *horizontal assessments* that focus on the efficiency criterion across several comparable interventions.

- Second, since some methods described in this report are far from being fully developed, considerable potential exists in their further development and, possibly, also in developing entirely new approaches.

In our understanding, even if frequency and quality of efficiency analysis are increased in the suggested way, it will not entirely satisfy expectations reflected in evaluation guidelines and national and multilateral policy documents. We therefore also recommend clarifying and specifying expectations in two ways:

- Third, expectations regarding efficiency analysis need to be adapted to what present and near-future methodology can realistically accomplish. This does not necessarily imply lowering expectations but rather clearly specifying the purpose for conducting efficiency analyses. The analysis levels introduced in this report allow for such a specification.

  For projects and simple programmes, we estimate that some level 2 and level 1 analyses should always be possible. This implies that the efficiency of several alternatives can be compared to each other and that efficiency improvement potential within specific alternatives can be identified.

- For more aggregated aid modalities such as complex programmes or budget support, we consider that efficiency assessment is usually limited to level 1 analysis. This implies that for these types of aid, the expectation of selecting the most efficient option by means of efficiency analysis alone, as for example in aid modality comparisons, needs to be reconsidered. For these types of aid, efficiency analysis is realistically restricted to identifying operational improvement potentials.

- Fourth, efficiency analysis should not be conducted whenever it is analytically possible. Instead, we recommend choosing carefully when to apply it.

Efficiency analysis itself also produces costs and benefits. Costs are usually resource and time investments for conducting the analysis. Benefits are, ultimately, increased development impact that can be reached in many ways, for example by providing assistance for the selection of more efficient interventions, by directly improving the efficiency of ongoing or planned interventions, by fostering learning through publication and dissemination of appraisal, evaluation or research reports, or by developing required skills through capacity development measures.

Depending on circumstances, the benefits of efficiency analysis may not justify its costs. Examples are expert judgements with low credibility, level 2 analyses without influence on the selection of interventions or efficiency analyses of interventions that are already known to have either very high or very low efficiency.

In such cases, the best use of resources may be to conduct a more efficient type of efficiency analysis or no efficiency analysis at all.

1. Introduction

1.1. Motivation for this Study

The efficient use of resources for development purposes is a compelling and internationally accepted principle. In order to maximise social and private benefits of aid interventions in developing countries, powerful analytical tools are needed.

When planning aid interventions, expected welfare effects can be appraised by analysing resource requirements and the results that are produced with these resources. If net welfare effects are positive, an intervention can be considered beneficial to society. If, in addition, resources needed to produce these effects are scarce, a rational choice is to select those interventions that produce the largest welfare effects based on those resources. In both instances, efficiency analysis is needed as a basis for decision-making.

Once aid interventions are under way or have ended, efficiency analysis can be conducted using empirical evidence. Hypotheses and assumptions made in appraisals can be tested and knowledge about what works and what does not work can be accumulated. Ex-post efficiency analyses can also serve as accountability mechanisms.

Apart from assessing the efficiency of entire aid interventions, more modest analyses may be helpful as well, for example assessing an isolated aspect that is linked to overall efficiency. In this way, the potential for operational improvements within a specific aid intervention can be identified without conducting broader analyses.

Because of its usefulness in decision-making and for learning, accountability and operational improvements, efficiency analysis has become a standard criterion for appraisal and evaluation of aid interventions. Many evaluation manuals and guidelines list efficiency among the standard evaluation criteria as, for example, the OECD DAC Criteria for Evaluating Development Assistance.[3]

In addition, efficiency analysis of public expenditures is advised in a more general way. In 1993, the United States White House issued an executive order that mandated federal agencies to conduct Cost-Benefit Analysis (CBA) for significant regulatory actions.[4] In deciding whether and how to regulate, agencies are required to assess all costs and benefits of available regulatory alternatives, including the alternative of not regulating. The definition of costs and benefits includes social benefits and costs that may be difficult to quantify.

Similarly, the German federal budget code explicitly prescribes the assessment of efficiency for all measures with financial effects on the federal budget, which includes most international aid. It also describes a number of methods that can be applied.[5] Other countries have similar provisions in place.

This high importance of efficiency analysis stands, however, in contrast to the frequency and quality with which it is applied in appraisals and evaluations of aid interventions.

---

[3] See, for example, OECD DAC (2010): Evaluating Development Co-Operation: Summary of Key Norms and Standards. Paris: OECD.

[4] The White House (1993): Executive Order 12866 of 30 September. Federal Register, Regulatory Planning and Review 58.

[5] BMF (2001): Allgemeine Verwaltungsvorschriften (VV-BHO), VV-BHO §7, H 05 01. Bundeshaushaltsordnung (BHO): BMF.

While the exact extent of this coverage remains unknown, several studies have shown that in those cases investigated, efficiency is analysed with low frequency and quality:

- An assessment of the quality and the management of 25 UNDP country-level evaluations conducted by the Swedish Agency for Development Evaluation (SADEV) finds that the efficiency criterion was dealt with in a *good or fair way* in only 20 per cent of the assessed evaluations, while the analysis was *poor* or *very poor* in 40 per cent and was missing altogether in the remaining 40 per cent.[6] The SADEV study authors conclude:

  > *Overall, the sample reports do not provide accurate assessments regarding efficiency. This is an area that needs attention as most terms of reference require information in this regard.*

- Another assessment of 34 project, programme and policy evaluations and organisational assessments conducted by Sida's Department for Evaluation (UTV) comes to a similar conclusion: only 20 per cent of the reports dealt with the efficiency criterion in a *minimally adequate manner* or better, while 65 per cent of the reports dealt with efficiency in a *not quite adequate manner*, had *significant problems*, or were rated *very poor*. Efficiency was not dealt with at all in 15 per cent of the cases.[7] The authors conclude:

  > *For a donor like Sida, questions about efficiency – broadly speaking value for money – are almost always likely to be of interest and relevance and, not surprisingly, most of the TOR in our sample included such questions. In most of the reports, however, the assessment of efficiency was technically quite weak. While all the reports included information about the resources spent on the intervention, very few provided a systematic assessment of the value of the benefits (outputs, outcomes, impacts) of the evaluated intervention in relation to the costs of producing them.*

- The World Bank Independent Evaluation Group (IEG) has conducted an assessment of 59 Global and Regional Partnership Program (GRPP) evaluations.[8] In roughly half (49 per cent) of the cases, efficiency was *not evaluated at all* and in a further 36 per cent, efficiency was *mentioned but not analysed in any meaningful way*. A *modest or substantial assessment of efficiency* was found in only 15 per cent of all cases.

- An analysis of the use of Cost-Effectiveness Analysis in evaluations commissioned by the European Commission screened 604 evaluation reports and impact assessments, of which only 161 (27 per cent) contained a substantial section or a chapter on efficiency.[9] Among other approaches to assess efficiency, the report identified 23 cases

---

[6] Hedblom; Hildenwall (2008): Strengthening the Results-Orientation in Sweden's Engagement in Multilateral Development Cooperation: An Evaluation of UNDP's Country Level Evaluation Activities. Karlstad: Swedish Agency for Development Evaluation.

[7] Forss; Vedung; Kruse; Mwaiselage; Nilsdotter (2008): Are Sida Evaluations Good Enough? An Assessment of 34 Evaluation Reports. Sida Studies in Evaluation. Stockholm: Swedish International Development Cooperation Agency.

[8] Statistical analysis of the coverage of the efficiency criterion in evaluations of Global and Regional Partnership Programs (GRPPs) prepared in 2007 by the World Bank Independent Evaluation Group (currently being updated). Internal document.

[9] Eureval-3CE; DG Budget (2006): Study on the Use of Cost-Effectiveness Analysis in EC's Evaluations. Lyon: Centre for European Expertise and Evaluation, pp. 10.

where a Cost-Effectiveness Analysis was conducted and only deemed 11 cases to fulfil basic quality criteria.[10]

In addition, the first three reports come to the conclusion that the efficiency criterion is addressed less frequently and with poorer quality than other evaluation criteria.

This fact is not entirely surprising since efficiency assessments usually build on other evaluation analyses, such as evaluations of effectiveness and cost analysis. Assessments of efficiency, therefore, are unlikely to exceed those analyses in frequency and quality. In addition, some methods for analysing efficiency require advanced expert methodology which may not be available.

Another recent IEG study[11] investigated the use of Cost-Benefit Analysis (CBA) in World Bank project appraisals and finds that both the frequency and the quality of CBAs have decreased over time, more than can be explained by a shift to less CBA-prone sectors alone. The frequency with which CBA has been applied in project appraisals has fallen from about 70 per cent of all projects in the 1970s to about 30 per cent in the early 2000s and in recent years. This is somewhat surprising, considering that World Bank policy mandates economic analysis, including CBA, for all investment projects.[12] In terms of quality, the share of projects with acceptable or good economic analysis in appraisal documents has declined from 70 per cent in a 1990 assessment to 54 per cent in a similar assessment conducted in 2008.

On the basis of existing policies, the above study results, and the feedback received during our interviews, we conclude that for many appraisals and evaluations of aid interventions, substantial gaps exist between what is required or intended, on the one hand, and what is delivered in terms of efficiency analysis, on the other.

The principal aim of this study is to provide assistance in filling such gaps by clarifying the concept, the analytic power and the limitations of efficiency analysis and by making available a catalogue of tools and methods for assessing aid efficiency. This information should have two effects:

- To help to increase the usefulness of efficiency analysis by alerting professionals to the existence and the potential of the full breadth of available methodology and to inform decision-makers as to the requisite outlays that will be required to conduct the work that is needed; and

- To help to design realistic policy and establish realistic professional expectations of what efficiency analysis can and cannot achieve.

---

[10] Strong analysis of cost, good enough analysis of effect(s), and comparison involving cost and effect.

[11] See World Bank Independent Evaluation Group (2010): Cost-Benefit Analysis in World Bank Projects. Washington: World Bank (received from IEG task manager Andrew Warner on 28 May 2010; unpublished at that time) and World Bank Independent Evaluation Group (2009): Annual Review of Development Effectiveness 2009: Achieving Sustainable Development. Washington: World Bank, pp. 43.

[12] "*The Bank evaluates investment projects to ensure that they promote the development goals of the borrower country. For every investment project, Bank staff conduct economic analysis to determine whether the project creates more net benefits to the economy than other mutually exclusive options for the use of the resources in question.[2]*

    [1] "Bank" includes IBRD and IDA, and "loans" includes IDA credits and IDA grants.

    [2] All flows are measured in terms of opportunity costs and benefits, using "shadow prices," and after adjustments for inflation."

Cited from: World Bank (1994): The World Bank Operational Manual, Operational Policies, Economic Evaluation of Investment Operations. Washington: World Bank, p. 1.

This study has a methodological research focus and does not attempt to produce judgements on the performance of development agencies or other groups in assessing aid efficiency.

This report summarises the findings of this study. Its target audiences are evaluation specialists and methodology researchers, but also the larger group of professionals in the field of aid evaluation without in-depth knowledge of methodology.

With the more general audience in mind, we have attempted to provide as much transparent explanation as possible. However, the descriptions in this report do not replace comprehensive textbook treatment of some well-documented efficiency analysis methods, such as Cost-Benefit Analysis, Cost-Effectiveness Analysis, Cost-Utility Analysis and methods for Multi-Attribute Decision-Making.

In addition, we have included aspects we felt were not well understood or controversial. This sometimes entails going into considerable detail that some readers may consider academic. Examples are the detailed treatments of the definition of efficiency in the second chapter and the discussion of specific aspects of Cost-Benefit or Cost-Effectiveness Analysis in the fourth chapter.

1.2. Study Approach

The study followed a simple approach of collecting and assessing tools and methods for the assessment of aid efficiency.

The collection and the initial assessment of methodology were based on a desk study and expert interviews. During the desk study, textbooks, journal articles, meta-studies on efficiency and evaluation manuals and guidelines were reviewed. References to these materials are inserted as footnotes directly in the text throughout this report. In addition, about one hundred evaluation reports from different aid agencies were screened opportunistically, i.e. using keyword-based searches with the aim of identifying methodology rather than assessing the degree to which reports covered certain methodologies. Among others, reports from Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ), the International Fund for Agricultural Research (IFAD), KfW Entwicklungsbank (KfW) and the World Bank were reviewed.

Based on the results of the desk study, 44 academics and practitioners with relevant experience in efficiency analysis were interviewed. In most cases, interviews were conducted in person during visits to evaluation units at the European Commission, the International Fund for Agricultural Development (IFAD), the Millennium Challenge Corporation (MCC), the main German implementing agencies (GTZ, KfW and InWEnt), and the World Bank (Independent Evaluation Group, Quality Assurance Group and World Bank Institute). Interviewees were primarily identified for their expertise and experience with efficiency analysis and care was taken to include different schools of thought. In several cases, interviews were complemented by substantial follow-up communication.

In addition to these interviews, a survey was conducted with 29 members of the OECD DAC Network on Development Evaluation.

The assessment of the methods and tools identified drew both on the desk study results and the interviews, and both sources were used to test our own understanding and obtain feedback on our conclusions.

1.3. Structure of this Report

The next chapter introduces the main concepts and clarifies efficiency-related terminology. Chapter 3 describes the dimensions along which methods and tools are characterised. These tools and methods are then described and assessed in chapter 4. The report concludes with a synthesis and several observations and recommendations of a more general nature.

Bibliographical references are provided as footnotes throughout the report and, additionally, in Annex B. We use the term evaluation for both ex-ante appraisals and evaluations that are conducted during, immediately after, or some years after an intervention. Similarly, we call the person conducting an appraisal or an evaluation simply an evaluator.

2. Key Concepts and Clarification of Terminology

Efficiency is a common term in everyday language, as well as in the technical language of several fields of science and business. This is both an advantage and a challenge. While most people easily relate to the general concept of efficiency, their specific understanding varies. First of all, therefore, we will introduce several concepts related to efficiency and clarify our terminology.

The Merriam-Webster Online Dictionary,[13] when elaborating on the subtle differences between the closely related words *effective* and *efficient*, explains:

> **"Efficient"** *suggests an acting or a potential for action or use in such a way as to avoid loss or waste of energy in effecting, producing, or functioning.*

The same dictionary explains the primary meaning of the noun *efficiency* as:

> **Efficiency** *[is] the quality or degree of being efficient.*

The concept of efficiency is applied in many ways. It can be applied whenever an input, however defined, is transformed into some result. And it can be applied to things, processes and systems.

In everyday life, we have many encounters with efficiency of objects or devices. The *fuel efficiency* of a car relates the fuel input to the mileage the car can generate with it. *Thermal fuel efficiency* is the amount of heat that can be created by burning a certain amount of fuel. *Energy efficiency* of a house describes the energy requirements for heating.

Then there is efficiency of processes. The *efficiency of a chemical reaction* is, for example, measured by the amount of substance produced in relation to the amounts of input reactants. The *production efficiency* of an assembly line of a car manufacturer essentially measures the efficiency of the production process, i.e. the number of cars produced per time or resource input.

Finally, systems are characterised by efficiency as well. When economists speak of *efficient markets*, they often refer to perfectly competitive markets that are not marred by real-world imperfections. The efficiency of a real market is then understood as the degree to which that market adheres to perfect competition rules, e.g. infinite amounts of willing, able and perfectly informed buyers and sellers, absence of entry and exit barriers, and so on. This type of efficiency is somewhat different from the simple input-results concept. The closest we may

---

[13] www.merriam-webster.com, visited on 12 July 2010.

get to that concept is by saying that an efficient system allows for efficient transaction. In the case of an efficient market, this would signify producer-consumer transactions that maximise the sum of producer and consumer surplus.

An important observation throughout these examples is that we are usually interested in comparing efficiency. Stand-alone efficiency information for a device, a process or a system is not very useful.

> *The fact that your car consumes eight litres of gasoline for every 100 kilometres – as a stand-alone information – may help you figure out whether you will have to stop for gas when driving home or for selecting a travel destination under a tight fuel budget. It, however, does not give you a feeling for whether eight litres per 100 kilometres is a good or bad value and it doesn't help you determine whether you should have bought this car in the first place.*

> *The situation changes dramatically if some extra efficiency information becomes available. Knowing, for example, the fuel-mileage ratio your car is supposed to have will allow you to decide whether to have your car's engine checked or whether it is running just fine. If you happen to know fuel-mileage ratios for other cars in the same class, you know whether to look proudly out of the window or to duck behind the wheel next time you pass environmental activists. It will also help you calculate your fuel savings compared to car alternatives and assist you in deciding when to scrap this car, whether to buy a new car at all, and if yes, what car to buy.*

In addition, it may be useful to investigate ways to improve efficiency. Comparisons may be helpful here as well but are not always necessary: a good mechanic can tweak an engine into higher performance after careful examination, even without access to other benchmarks.

Another important observation is that efficiency is described in two different ways:

- On the one hand, efficiency is described as the *conversion* or *transformation* of something into something else, for example of fuel into heat or, more generally, of some input into some result. We refer to this as *transformation efficiency*: the degree to which inputs are used to produce results determines efficiency.

- On the other hand, efficiency is also described by means of a single quantity or rule, for example in the cases of net welfare effects or the degree to which a market fulfils perfect market rules. The larger this quantity, the more efficient the system (or the intervention) is considered to be. We refer to this type of efficiency as *optimisation* efficiency: the degree to which an optimisation rule is fulfilled indicates efficiency.

In the language of welfare economics, transformation efficiency can for example be measured by benefit-cost ratios while optimisation efficiency is measured by net benefits.

The distinction between these two efficiency concepts may appear superfluous at first: aren't they just two equivalent measures for the same thing?

As we will see later in this report, they are not. Choosing the most efficient development interventions based on transformation efficiency information may lead to different results than selecting interventions based on their optimisation efficiency.[14]

---

[14] See the discussion in section 2.2.4 and the more detailed treatment in the context of Cost-Benefit Analysis and Cost-Effectiveness Analysis in sections 4.3.2 and 4.3.4 respectively.

In the next sections, we will discuss how efficiency is understood in international development and in welfare economics, clarify terminology and provide explanations that will be needed later in this report.

We begin by presenting the closely related definitions of efficiency by *inputs* and *results* in section 2.1 and based on *welfare economics' costs and benefits* in section 2.2. Sections 2.3 and 2.4 briefly introduce challenges and approaches for assessing effects and costs, which are the two types of information required for most methods for assessing efficiency. In section 2.5, two special efficiency considerations, *cost minimisation* and *yield maximisation*, are placed into the context of general efficiency analysis. Section 2.6 highlights the importance of efficiency analysis as a basis for rational decision-making and for welfare optimisation. Finally, section 2.7 explains the necessity and the consequences of choosing a particular analysis perspective when conducting efficiency analyses.

2.1. Definition of Efficiency by Transformation of Inputs into Results

The OECD Development Assistance Committee defines efficiency as follows:

> *Efficiency is a measure of how economically resources/inputs (funds, expertise, time, etc.) are converted to results.*[15]

In this definition, the term *results* stands for the output, outcome or impact of a development intervention. Any of these quantities can be intended or unintended as well as positive or negative. In addition, inputs and results need to be causally related to the intervention, as will be explained in more detail later. An intervention is considered *economical* if the costs of scarce resources used approximate the minimum needed to achieve planned objectives.[16]

This definition explains efficiency *as transformation efficiency*: inputs are transformed into results and their relative amounts represent an efficiency measure.

In a survey we conducted with 29 representatives in the *OECD DAC Network on Development Evaluation*, the majority of that group (55 per cent) indicated that their home organisations worked with this definition of efficiency, while the remainder (45 per cent) indicated that their definition described the same concept but had a different wording.[17]

In a slightly different vein, the OECD DAC describes the *evaluation criterion* efficiency as follows:

> *Efficiency measures the outputs – qualitative and quantitative – in relation to the inputs. It is an economic term which is used to assess the extent to which 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.*[18]

Both definitions introduce efficiency by relating effects to the inputs needed to produce them. The second definition limits the assessment of efficiency to the level of outputs (or remains

---

[15] OECD DAC (2002): OECD DAC Glossary of Key Terms in Evaluation and Results-Based Management. Paris: OECD Development Assistance Committee.

[16] Explanations based on ibid.

[17] Survey participants were not asked to comment on the second definition presented in this section.

[18] OECD DAC (2010): Evaluating Development Co-Operation: Summary of Key Norms and Standards. Paris: OECD.

ambiguous about this by using the term "results" in the explanatory text) whereas the first definition covers outcomes and impacts as well.

For the purposes of this report, we will follow the first definition since it is more general and contains the second definition as a special case.[19]

If outputs are assessed in relation to inputs, this is sometimes referred to as *production efficiency*. If, instead, higher-level results are analysed in relation to inputs, the term *allocation efficiency* is used.

This definition pretty much predetermines which methods and tools this report covers. All methodology that relates the results of an aid intervention to the inputs[20] of that intervention is of principal interest.

However, this definition does not imply the following:

- We have found some publications that struggle to differentiate "efficiency analysis" from "analysis using methodology from the field of economics". In our opinion, this may simply be due to using the word *economical* as a synonym for *efficient* and subsequently jumping to the conclusion that *analysing efficiency* is equivalent to *applying economic analysis*.[21]

  The OECD DAC definition and this report do not preselect methods by the professional field they originate from. Some, but not all, economic analyses qualify as efficiency analyses. Similarly, some non-economic methods qualify while others do not.

- Several people interviewed for this report asked whether this report was about methods for the assessment of operational efficiency of aid agencies or the development efficiency of aid interventions.

  While it is true that institutional efficiency of aid agencies is usually not assessed in conjunction with evaluations of their projects or programmes, we see no need to reduce the scope to one or the other. In our understanding, an assessment of an aid intervention's efficiency automatically includes, among many others, questions about the implementing agency's operational efficiency if it is conducted from a perspective that includes the agency.

2.2. Definition of Efficiency in Welfare Economics

Welfare and development economists think about welfare from the perspective of an entire society. Interventions that take place within the socio-economic system that constitutes a society are likely to somehow affect overall welfare. Welfare economists aim at understand-

---

[19] This choice is shared by other professionals as well. We asked 28 representatives in the OECD DAC Network on Development Evaluation how much of an interventions results chain should be covered by efficiency analysis in order to be satisfactory. Only 15 per cent voted for costs and outputs (and none for costs only), while 82 per cent felt that outcomes should be included (32 per cent voted for inclusion of costs, outputs and outcomes and 50 per cent for inclusion of the entire results chain; 1 respondent gave an answer that didn't fit into any category).

[20] From this point on, we understand resources to be part of input.

[21] The same is true in German language where "Wirtschaftlichkeit" (economy) is sometimes used as synonym for "Effizienz" (efficiency). The resulting confusion is addressed, for example, in BMZ (2001): Hauptbericht über die Evaluierung: Berücksichtigung von Wirtschaftlichkeitsüberlegungen bei EZ-Vorhaben. Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung.

ing and estimating welfare effects of different interventions, which is, fundamentally, an efficiency consideration.

Based on individual welfare considerations, two principal efficiency concepts of welfare economics – that of *Pareto* and of *Kaldor-Hicks efficiency* – can be introduced. We will present both concepts in the next section.

If, instead, we are interested in directly addressing the welfare of an entire society (and not only that of individuals), we have two options. On the one hand, individual welfare contributions can be aggregated as discussed in detail in section 2.2.2. On the other hand, overall welfare can be directly defined and described by a social welfare function as summarised in section 2.2.3.

Rather than talking about welfare improvements, welfare economists often work with generalised benefits and costs that are introduced section 2.2.4.

We conclude by noting that, while usually treated quantitatively, the principles discussed in this section do not require quantification (section 2.2.5).

2.2.1. Pareto and Kaldor-Hicks Efficiency

If a society undergoes changes, these can produce gains and losses for each individual. If such changes produce winners but no losers, we can be sure that society is better off after the change.[22] Such changes are called *Pareto improvements* and can be defined as follows:

> *Pareto improvements to a socio-economic system make at least one person better off and no one worse off.*

After each Pareto improvement, society is better off than before. If we have applied all possible Pareto improvements, we have maximised society's welfare under the Pareto rule. Such a socio-economic system is then considered *Pareto efficient*.

This concept describes efficiency as *optimisation efficiency*: people are made "better off" and the optimal, efficient state is reached when all such improvements have been exhausted. In contrast to transformation efficiency, it remains unclear what is actually used as input to produce this result.

It should also be noted that, in order to make these statements, we do not have to compare the welfare changes of one individual to those of another. In technical terms, the state after a Pareto improvement *dominates*[23] or *outranks*[24] the state before. As we will discuss in more detail in the next section, comparisons of individual welfare effects entail some intrinsic difficulties.

However, Pareto improvements only exploit a small fraction of interventions that can be considered beneficial to society. Matthew Adler and Eric Posner describe this as follows:

> *It is likely that the Pareto standard would reject desirable projects that would be approved under an uncontroversial welfare function. For example, a vaccine that improved the health of millions but required a tax of one dollar on someone unaf-*

---

[22] Strictly speaking, this is only true if differences in individual welfare have no influence on welfare.

[23] See, for example, Keeney; Raiffa (1993): Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge England; New York, NY, USA: Cambridge University Press, pp. 69.

[24] See, for example, Eureval-3CE; DG Budget (2006): Study on the Use of Cost-Effectiveness Analysis in EC's Evaluations. Lyon: Centre for European Expertise and Evaluation, pp. 6.

fected (who is not altruistic) would violate the Pareto standard but surely is morally required or at least permissible.[25]

Edward Gramlich[26] explains this more generally by the fact that the Pareto rule usually avoids *type II errors* (where a policy is not beneficial but approved by decision-makers) but will generate many *type I errors* (where a policy is beneficial but not approved by decision-makers).

What is needed is a less demanding screen that allows for winners and losers but still ensures that overall gains to society are positive. The *Kaldor-Hicks rule*, for example, builds on the Pareto rule but, in addition, allows for trade-offs.

> *Kaldor-Hicks improvements to a socio-economic system make the gainers so much better off that these, theoretically, could compensate the losers and still be better off than before. Therefore, these improvements are sometimes also called potential Pareto improvements, even if actual compensation might be impossible.*

If no more Kaldor-Hicks improvements are possible, the socio-economic system is considered *Kaldor-Hicks efficient*. The Kaldor-Hicks rule allows for a much wider range of interventions because it includes a special type of trade-off.

As with the Pareto rule, the Kaldor-Hicks rule introduces efficiency as *optimisation efficiency*, i.e. the degree to which the best possible state is reached when applying successive Kaldor-Hicks improvements.

Whether a Kaldor-Hicks efficient state also represents a state of maximal welfare depends on the specific welfare model used.[27] One solution is to turn Kaldor-Hicks improvements into Pareto improvements by actually executing compensations. These are, however, not easy to quantify since winners may have incentives to hide their gains and losers may have incentives to exaggerate their losses. If compensations are executed financially, several tax schemes have been devised for this purpose.

An important general consideration when allowing for trade-offs – as part of the Kaldor-Hicks rule or any other non-Pareto efficiency concept – is that *distributional effects* may become

---

[25] Cited from: Adler; Posner (2006): New Foundations of Cost-Benefit Analysis. Cambridge: Harvard University Press, p. 20.

[26] See: Gramlich (1990): A Guide to Benefit-Cost Analysis. Englewood Cliffs, N.J.: Prentice-Hall, p. 32.

[27] The Kaldor-Hicks rule does not imply quantifying and comparing individual welfare contributions but instead values welfare on an ordinal scale from strictly individual perspectives. Depending on the way individual welfare is quantified and society's welfare is aggregated from these individual welfare contributions, Kaldor-Hicks improvements may represent an increment or decrease in society's welfare. Below, we provide a simple example where a Kaldor-Hicks improvement does not lead to an improvement in total welfare:

Assume that Sara is a fruit lover while John is not so keen on fruit. Sara's and John's preferences for fruit are described by their willingness to pay for apples and oranges as indicated below. Let's make the simplifying assumption that these numbers are constant.

	Sarah	John
Willingness to pay for an apple	2 euros	50 cents
Willingness to pay for an orange	5 euros	30 cents

If Sara gives an apple to John, and John gives an orange to Sara, everybody is better off; the fruit exchange is a Pareto improvement. If, instead, we only conduct the first part of this transaction (Sara gives her apple to John), we have a Kaldor-Hicks improvement. John could compensate Sara by giving her his orange – or a part of that orange – which would compensate Sara for her loss of an apple. If we determine total welfare changes by simply adding up individual welfare changes, society's welfare has increased by 3.20 euros in the first case but has decreased by 1.50 euros in the second case.

invisible in an analysis that focuses exclusively on welfare changes for the entire society. For example, if losses by one group are balanced by another group's gains, net changes to society's welfare are zero while the two groups have experienced welfare changes.

Obviously, a solution to this is to simply conduct the analysis for all sub-groups separately as well as for society as a whole. In this way, distributional effects are rendered transparent.

2.2.2. Measuring and Aggregating Individual Welfare

The Pareto and Kaldor-Hicks rules are special in the sense that they do not require information on how individual welfare is actually defined or determined and on how the welfare of a society can be quantified.

If, instead, we aim at quantifying welfare changes to society, we have two principal options:

- On the one hand, we can aggregate society's welfare from its individual components as described further below in this section. This is a microscopic model in the sense that a macroscopic quantity (society's welfare) is explained by microscopic contributions (the welfare of individuals).

- On the other hand, we can directly define social welfare functions that depend only on aggregated variables. This is a macroscopic model in the sense that a macroscopic quantity is related to other macroscopic quantities. Obviously, mixed approaches are possible as well. Social welfare functions will be described in section 2.2.3.

The problem of developing a microscopic model for society's welfare can be broken down naturally into three related aspects:

- Definition of individual welfare;

- Quantification of individual welfare; and

- Aggregation of individual welfare contributions.

Jean Drèze and Nicholas Stern comment in a very similar vein (combining the first and second point under "assessment" and using households as minimal units):

> *The evaluation of social welfare is naturally based (at least partly) on assessments of the well-being of individual households, supplemented by interpersonal comparisons of well-being. The latter are embodied in what we shall call "welfare weights".[28]*

In the following paragraphs, we discuss each of these aspects separately.

Definition of individual welfare. The Pareto and Kaldor-Hicks efficiency concepts use the term "better off" to define improvements in individual welfare. But what exactly does "better off" imply? What constitutes individual welfare improvements?

These questions touch on a rich and established field of welfare theorists' research which we do not attempt to summarise comprehensively. We will, however, describe some key points.

Most welfare economists think about individual benefits and costs in terms of the *satisfaction of personal preferences*. This approach is based on the important assumption that individual welfare is essentially responsive to the individual's point of view.

---

[28] Dreze; Stern (1994): Shadow Prices and Markets: Policy Reform. Cost-Benefit Analysis. Layard and Glaister, pp. 61.

Other views exist as well.[29] Some economists, for example, conceptualise individual welfare as pains and pleasures and focus on *subjective well-being* and *happiness* that is mapped to specific mental states. This welfare account shares subjectivity with the preference satisfaction view. Other, more objective, concepts are based on lists of goods and values that represent different dimensions of human welfare. Other authors feel that neither welfare account is satisfactory on a stand-alone basis. For example, Matthew Adler and Eric Posner[30] illustrate some circumstances where the preferences-only approach may lead to results many people intuitively or morally disagree with. One example is that of *non-ideal preferences*, i.e. preferences that violate moral standards or are unlawful. Another example is that of *disinterested preferences*, i.e. preferences entirely driven by moral standards or causally disconnected from real events, for example if the person expressing his or her preference will never learn about the outcome that preference is based upon.

These deliberations on different individual welfare concepts may remain somewhat philosophical in the sense that they are unlikely to be easily corroborated by empirical evidence. Nevertheless, they lay the foundations of what is measured when conducting efficiency analysis. For the purposes of this report, it is important to keep in mind that different conceptual models exist, each based on a specific view of individual welfare.

Most methods for assessing efficiency portrayed in this report focus on the question of *how* something is measured and not so much on *what* is measured, although some methods implicitly build on specific conceptual welfare models. In all cases, individual perceptions on welfare of all persons participating in an evaluation of efficiency, e.g. the evaluator and contributing stakeholders, influence the analysis result. If the conceptual welfare models (*what* is measured) of those people were to change, results obtained with efficiency analysis methods would change as well and some methods would need to be adapted.

Quantification of individual welfare. Based on the often used preference-satisfaction approach, several methods have been developed to quantify individual welfare, each with its own specific strengths and weaknesses.

A common approach is to approximate the strengths of preferences by *willingness to pay*, i.e. the amount of money a person would be willing to pay for his or her preference satisfaction. Individual willingness to pay can either be directly measured or is induced from further analysis. As a measure, willingness to pay has clear advantages. It indicates the intensity of preferences on a quantitative cardinal scale that allows comparing and aggregating individual preferences. A potential complication lies in the fact that willingness to pay not only measures the intensity of a person's preferences for specific outcomes but also how that person values money, i.e. the person's preference for money. The latter, in turn, is likely to depend on a person's wealth.

Other approaches do not build on willingness to pay but convert ordinal survey results into abstract quantitative values that reflect the intensity of preferences, often referred to as *utility*. This has the advantage of allowing calculus to be applied, for example when combining one person's preferences for different attributes into a single utility number. Obviously, it also requires assumptions that need to be kept in mind when interpreting results, as illustrated for the standard problem of converting ordinal survey results into cardinal values in the following example:

---

[29] For an overview, see, for example, Adler; Posner (2006): New Foundations of Cost-Benefit Analysis. Cambridge: Harvard University Press, pp. 28.
[30] Ibid. pp. 33.

Imagine assigning utility values -2, -1, 0, 1 and 2 to the qualitative statements "very unfavourable", "unfavourable", "neutral", "favourable" and "very favourable", respectively. This implies two assumptions. First, we assume that all intervals between the ratings are equal and, second, we have associated "neutral" with the zero point of that scale. None of these assumptions are written in stone. More detailed investigation could, for example, reveal that a survey participant feels that "favourable" and "very favourable" lie much closer together than "neutral" and "favourable" or that even "very unfavourable" outcomes are better than nothing, placing the zero point somewhere to the left of that qualitative rating.

If utilities along several attributes are compared and aggregated, additional normalization and relative weighting techniques are required that imply further assumptions.

The most general and, at the same time, most rigorous treatment of individual preference satisfaction is condensed in the *theory of cardinal utilities*.[31] In utility theory, cardinal utilities are derived and measured by a mental lottery often referred to as *standard gamble*[32] that will be explained in more detail when describing approaches for Multi-Attribute Decision-Making (MADM) in section 4.3.1. Utility theory is also able to deal with multivariate and interdependent utility functions. In practical applications, individual preferences are often approximated by group averages and further approximations are made when estimating these average quantities. Individual willingness to pay is, for example, sometimes approximated by market prices for goods or services.

All of these approaches are sometimes criticised because they are not perfect. Individuals need to understand potential consequences upfront in order to value different outcomes appropriately and any method makes a number of assumptions. However, if limitations are kept in mind, these approaches provide a powerful set of tools to quantify the intensity of individual preferences.

Aggregation of individual welfare contributions. Until now, we have only considered defining and quantifying welfare for individuals. In order to arrive at a description of society's welfare, individual welfare contributions need to be aggregated. The obvious alternative to this is to directly define welfare on the level of entire societies by social welfare functions as described in section 2.2.3.

For descriptions of welfare based on preference satisfaction, the aggregation of individual welfare contributions leads to the *problem of interpersonal comparisons of preferences*. Each person can express his or her own preferences for different outcomes but how can we compare the intensity of preferences of different individuals?

Several approaches exist. First, we can introduce an absolute external reference. In the case of individual welfare measured by willingness to pay, this external reference would be money. This allows us to calculate overall welfare by simply adding up individual surpluses. Second, we can normalise individual preferences by placing them on the same cardinal scale. A typical approach is to set lowest preferences to zero and highest preferences to 1. However, several different approaches to normalisation exist as well.

---

[31] See, for example Keeney; Raiffa (1993): Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge England; New York, NY, USA: Cambridge University Press.

[32] See section 4.3.1 of this report for an example.

All of these approaches imply assumptions that need to be considered when interpreting results. As indicated earlier, willingness to pay also reflects peoples' state of wealth. If we wanted welfare to reflect the intensity of desires for certain outcomes only, we would need to somehow correct for this when using willingness to pay as the measurement unit. If, for example, we measured individual welfare by the amount of time people would be willing to invest in obtaining specific outcomes, we might obtain a different result when adding up individual contributions, reflecting availability of time rather than of money. Similarly, normalising preference scales based on highest and lowest preferences might obfuscate relative intensities of preferences.[33]

Of course, other mechanisms for combining individual preferences exist or can be devised, each with particular strengths and limitations. The important point is that – when aggregating individual welfare contributions – assumptions on interpersonal welfare comparisons need to be made. Kenneth Arrow formally made that point in 1951[34] when he showed that, under reasonable assumptions, without explicitly addressing the question of interpersonal comparison of preferences, there is no procedure for combining individual rankings into group rankings.[35]

Based on this type of consideration, some academics reject macroscopic theories of welfare built on microscopic welfare contributions that are each valued from the perspective of the individuals concerned. Others argue in favour of interpersonal welfare comparisons, such as Matthew Adler and Eric Posner[36] who feel that these academic concerns should not block addressing problems that have clear intuitive solutions as illustrated below:

> *In one world, Felix feels fine but thousands die painful premature deaths; in another world, the deaths and pain are averted but Felix has a headache. In one world, many workers lose their jobs, and suffer depression, family strife, or other ill effects of protracted unemployment; in another world, the unemployment doesn't occur but a small number of executives can now only afford BMWs rather than Lamborghinis. Although the two worlds in these scenarios are Pareto-noncomparable, the second world in each scenario is (intuitively) much better for overall welfare.*

In our opinion, the most convincing approach to dealing with interpersonal welfare comparisons is to bypass the problem. For example, Ralph Keeney and Howard Raiffa[37] argue that

---

[33] Consider two people, Sara and John, and their preferences for fruit. Let's say Sara's utility for having an apple is 0.5 and John's is 0.8. Both are measured on normalised scales, i.e. the fruits that they like least are reflected by zeros and the fruits they like best are reflected by ones. At this stage, no comparison is possible. All it tells us is that, on Sara's personal fruit preferences scale, apples rate rather low, i.e. that there are other fruits that rate higher. For John, instead, apples rate among the types of fruits he, relative to other fruits, likes most. If we calculated the group's utility by simply adding Sara's and John's individual utility functions, we would come to the conclusion that we create more preference satisfaction by giving an apple to John (+0.9 utility points) instead of to Sara (+0.2 utility points). Now imagine that Sara craves fruits in general and that John doesn't like fruit so much. Our common sense tells us to value Sara's preferences as stronger than those of John but we don't know by how much.

[34] Arrow (1951): Social Choice and Individual Values. New York: Wiley.

[35] See, for example, Keeney; Raiffa (1993): Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge, England; New York, NY, USA: Cambridge University Press, p. 524.

[36] Adler; Posner (2006): New Foundations of Cost-Benefit Analysis. Cambridge: Harvard University Press, p. 41.

[37] See, for example, Keeney; Raiffa (1993): Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge England; New York, NY, USA: Cambridge University Press, p. 515.

decisions are usually taken by individuals and not by groups. Therefore, devising ways to sum up the true, self-perceived preferences of different individuals is not needed. Instead, all the different individual preferences are assessed from the perspective and reference frame of a single individual, often referred to as the decision-maker.

This is the domain of decision theory and multi-attribute utility theory,[38] a well developed methodology for systematically assessing decision-making options from the perspective of such a decision-maker. From that perspective, preference satisfaction of other people and groups remains of importance. Consequently, the decision-maker will take individual welfare changes of others into account when taking decisions, just as he or she will take his or own direct preferences into account. Since everything is valued from the decision-maker's perspective, the problem of interpersonal comparison of preferences as described at the beginning of this section is shifted to the determination of the decision-maker's utility for the preferences he or she perceives in other people. This point of view provides the basis for the methods of Multi-Attribute Decision-Making (MADM) described in section 4.3.1 of this report.

2.2.3. Social Welfare Functions

Most practitioners will not go into the academic depths outlined in the previous section when describing social welfare but will rather postulate a *social welfare function*.

Social welfare functions describe a direct functional relationship between a measure of society's welfare and a number of input parameters. These input parameters can be individual (e.g. individual income) or societal (e.g. life expectancy, per capita income). In this way, a value for society's welfare can be computed for each state of society that is described by a full set of input parameters.

The advantages of social welfare functions are their transparency and usability. Definitions made are visible to everybody and, based on the model defined by a specific social welfare function, welfare for different states of society can easily be quantified.

Obviously, different social welfare functions correspond to different models for society's welfare and reflect different understandings of welfare, different assumptions about welfare drivers and different welfare weights for different outcomes. For example, the social welfare function defined by a foreign donor may look quite different from an analogous function defined by a beneficiary group.

2.2.4. Definition of Efficiency Based on Costs and Benefits

Rather than talking about overall welfare improvements, welfare economists often use the terms *benefits*, *costs* and *net benefits*. In analogy to microeconomic investment theory, net benefits are composed of *total benefits* and *total costs*. However, in contrast to microeconomics, benefits and costs are defined more broadly, reflecting the intention to grasp all relevant welfare changes:

- The *benefits* of an intervention are all private and social, direct and indirect, tangible and intangible effects of that intervention;

- The *costs*, defined in equally general terms, represent everything required to produce these effects.

---

[38] See, for example, ibid.

Both benefits and costs are net quantities in their own terms and may be composed of many different contributions that may be positive or negative.

Net benefits are a useful concept since they summarise total net welfare gains of an intervention and also provide an idea of how to actually calculate them.

A widely used efficiency analysis method in the field of economics is the Cost-Benefit Analysis (CBA), which is discussed in more detail in section 4.3.2. CBA estimates both costs and benefits and monetary units and is therefore capable of explicitly calculating net benefits.

Starting from benefits and costs, efficiency can be described both by using *optimisation* and by *transformation efficiency*:

- *Optimisation efficiency*. If efficiency is described by net welfare effects, for example by the net benefits caused by an intervention, the optimisation efficiency concept is used.

  If all interventions with positive net benefits have been applied to a socio-economic system, the system is efficient under the net benefit rule.

  If it is not possible to implement all interventions with positive net benefits because of constraints (e.g. limited financial resources), it may be sensible to select those with the highest net benefits in order to optimise welfare under these constraints.

  These net benefit rules and the use of these rules in decision-making are explained in more detail in section 2.5 of this report.

- *Transformation efficiency*. Based on costs and benefits we can measure efficiency by the transformation of costs into benefits.

  This interpretation is straightforward and often applied. It has the advantage that it allows us to calculate standardised efficiency measures, for example benefit-cost ratios, cost-effectiveness ratios, cost-utility ratios and Economic Rates of Return (ERRs).

  There are, however, also difficulties related to this definition. For example, if used for decision-making, selection of interventions based on how well costs are transformed into benefits may not optimise welfare, for example if scarce resources are not adequately reflected by costs and are, therefore, not optimally allocated. Problems can also occur if the definitions of costs and benefits are unclear, for example if some negative benefits are considered as costs instead. These and other issues are further elaborated in the context of Cost-Benefit Analysis and Cost-Effectiveness Analysis in sections 4.3.2 and 4.3.4, respectively.

  Alternative definitions of transformation efficiency are possible but are not well-documented.[39]

---

[39] One option is to consider how particular resources that are needed for an intervention are converted into net benefits. This definition is able to reproduce the results of decision-making based on the net benefit rules described in section 2.6 if implementation constraints can be expressed by scarce resources. Since it is based on net benefits, it is also immune to whether negative benefits are considered costs or not since these differences cancel out when the difference between benefits and costs is determined. However, they do not cancel out if costs and benefits are considered separately or as ratios. Another option that is discussed in the context of Cost-Effectiveness Analysis (section 4.3.4) is to consider how some resources are converted into *residual* net benefits that are defined as net benefits without these resources. This option has the mathematical advantage that it does not consider the same resources on both sides of the equation, i.e. as input and as part of results.

2.2.5. Efficiency as a Qualitative Concept?

In our observation, many experts feel that analysis of efficiency, specifically when based on benefits and costs, is essentially a quantitative process that is largely owned by development economists. Several experts also feel that this analysis domain has little place for qualitative evaluation techniques and some expressed their concern that analysts may willingly or unwillingly sacrifice realistic description of complexity for the sake of producing quantitative analysis results.

Similar concerns are reflected in several textbooks that advocate more participatory evaluation approaches that place increased importance on qualitative understanding rather than on quantitative measurement, quite contrary to the perceptions described above. In these approaches, the evaluator acts as moderator and catalyst and less as objective analyst. Pushing things even further, the *constructivist* approach to evaluation fundamentally questions the scientific paradigm underlying most evaluation approaches.[40]

However, while outspoken regarding the limitations of quantitative scientific measurement in general, these textbooks remain silent regarding how the efficiency concepts described in this chapter could be applied in alternative ways.

We feel that, even for approaches rejecting quantitative or other scientific evaluation approaches, the concept of efficiency remains viable. We do not consider efficiency analysis to be restricted to quantitative, scientific approaches. This is, for example, illustrated by approaches driven by stakeholders or fundamentally based on stakeholder feedback, as described sections 4.2.4 and 4.2.5 of this report.

2.3. Determining Effects Caused by an Intervention

For almost all[41] methods for assessing efficiency described in this report, information about the effects and the costs of an intervention is required in some way. In this section, we focus on how effects can be assessed and will discuss the assessment of costs in the next section.

Information on the effects of a development intervention can be gathered, analysed and synthesised by a host of methods. These cover the entire spectrum of qualitative, mixed and quantitative methods for assessing the *effectiveness* of development interventions and range from ethnologic to econometric approaches. As these approaches are described in many textbooks on evaluation and on specific data gathering and analysis methods, we will not describe them in more detail here. Instead, we highlight a crucial concept that needs to be observed for any solid assessment of efficiency.

When analysing interventions, we are only interested in those changes that are *caused* by the intervention. This is also often referred to as *attribution*.[42]

---

[40] The key differences between the constructivist and more traditional approaches to evaluation can be summarised as follows (adapted Guba; Lincoln (1989): Fourth Generation Evaluation. Newbury Park, Calif.: Sage Publications): The existence of multiple, socially constructed realities ungoverned by any natural laws, causal or otherwise (in contrast to one objective reality governed by cause and effect); Strong interlocking of the inquirer and the "inquired into" leading to dependence of inquiry results on the inquirer (in contrast to a detached observer); A methodology focusing on a joint construction of a case among all inquirers and respondents (instead of a scientific observation or measurement).

[41] Some methods, instead, assess efficiency directly without prior assessment of effects and costs. See, for example, the *Comparative rating of efficiency* approach described in section 4.2.4. This section's remarks regarding attribution in this section apply nevertheless.

Causally attributable changes can either be (mostly positive) effects that are driven by the intervention, or (mostly negative) effects that are avoided by the intervention. In both cases, other changes not caused by the intervention usually exist and often exceed the effects produced by the intervention in magnitude. These other changes, positive or negative, occur no matter whether the intervention takes place or not and are of no direct interest to us, unless needed to calculate causally related effects. In order to determine the total effects caused by an intervention, all changes in the situation with the intervention need to be subtracted from all changes in the situation without the intervention. This eliminates changes that are the same in both scenarios (the ones that do not interest us) and gives us the full amount of causally related effects.

In the case of several alternative interventions, each intervention can be compared to the scenario without an intervention. In some instances, for example if the scenario without an intervention is not a realistic alternative, one intervention can be chosen as a base scenario and all other alternative interventions can be assessed in relation to it.

As convincing and important as it is, the concept of reference scenarios remains a theoretical one. Since we cannot turn back time, we can never observe the exact same situation with and without an intervention or with an alternative intervention. Therefore, other ways need to be found to approximate what the situation without the intervention or with another intervention *would have been*.

A wealth of literature exists on this topic which we will not attempt to summarise in any detail, apart from the following remarks.

The most prominent way – and sometimes claimed to be the only way – in which causality can be inferred from observation, is that of the randomised experiment in which both "treatment" and "control" group participants are randomly selected from the same sample and are therefore statistically equal in all properties apart from the intervention. Such experiments can be conducted by observation in the real world (e.g. randomised controlled trials),[43] by simulation of the reference scenario (e.g. quasi-experiments), or can be approximated by before-after considerations without control groups if confounding effects are negligible.[44]

Since many evaluators keep a somewhat respectful distance to these methods or feel that other methods are more appropriate, in many instances, more accessible yet less reliable approaches are chosen that depend on the common sense of the evaluator and on the investigation of the theory of change of an intervention (e.g. outcome mapping,[45] contribution

---

[42] On causal inference see, for example, Pearl (2009): Causality: Models, Reasoning, and Inference. Cambridge; New York: Cambridge University Press and on experiments and quasi-experiments, for example, Shadish; Cook; Campbell (2002): Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin.

[43] See, for example, Duflo; Kremer (2003): Use of Randomization in the Evaluation of Development Effectiveness. World Bank Operations Evaluation Department Conference on Evaluation and Development Effectiveness. Washington, D.C. and other methodology documentation provided at www.povertyactionlab.org/methodology, including free training course material. Website visited on June 21, 2010.

[44] For a thorough overview see, for example, Shadish; Cook; Campbell (2002): Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin.

[45] For an introduction see, for example, Hearn; Jones (2009): Outcome Mapping: a Realistic Alternative for Planning, Monitoring and Evaluation. London: Overseas Development Institute, and other resources on www.outcomemapping.ca/resource/.

analysis[46]), are anecdotal (e.g. Most Significant Change[47]) or build on stakeholders' capacities in grasping and valuing causal relations (e.g. Utilization-Focused Evaluation[48]). Finally, in constructivism, the very idea of objective reality and causality is rejected and attribution is replaced by other concepts, for example that of *mutual simultaneous shaping* (e.g. Fourth Generation Evaluation[49]).

Unfortunately, in many appraisals and evaluations, attribution is considered implicitly or not at all when effectiveness is analysed and effects caused by an intervention cannot be distinguished from unrelated changes. Some of the experts interviewed for this report described the analysis of attribution as the single most challenging hurdle for reliable effectiveness and efficiency analysis.

Apart from these remarks on attribution, we will not stress the importance cause and effect anymore but will quietly assume that costs and effects as well as inputs and results have been attributed to the intervention at hand in some reliable way. While attribution, in our opinion, remains a matter of concern in many appraisals and evaluations, it is not specific to the analysis of efficiency but rather to evaluation in general and therefore has its place in standard textbooks on evaluation rather than in this report.

2.4. Determining Costs of an Intervention

In addition to information about effects, information about costs is needed for most efficiency-assessment methods.

Costs can be assessed in a number of ways.[50] Roughly, the following approaches exist:

- *Cost summary.* In this simple and entirely descriptive approach, the evaluator studies and synthesises financial information on actual expenditures from available documentation, for example from annual financial or audit reports or from bookkeeping databases. The cost synthesis often contains a table listing expenditures for several cost categories over several years and some explanatory text.

- More systematic approaches to cost analysis are the closely related *ingredients method* and the *resource cost model[51]* that systematically scan through and valuate different cost categories.

- *Financial feasibility assessment.* For project appraisals, an additional financial feasibility analysis can be conducted on the basis of a cost summary by comparing the fi-

---

[46] Mayne (2001): Adressing Attribution through Contribution Analysis: Using Performance Measures Sensibly. Canadian Journal of Program Evaluation.

[47] See, for example, Davies; Dart (2005): The 'Most Significant Change' (MSC) Technique: A Guide to Its Use Version 1.0 , available, for example, at http://www.mande.co.uk/docs/MSCGuide.htm (last visited on 26 July 2010).

[48] Patton (2008): Utilization-Focused Evaluation. Thousand Oaks: Sage Publications, pp. 492.

[49] Guba; Lincoln (1989): Fourth Generation Evaluation. Newbury Park, Calif.: Sage Publications, pp. 97.

[50] For descriptions of cost analysis approaches see, for example, Greenberg; Appenzeller (1998): Cost Analysis Step by Step : a How-to Guide for Planners and Providers of Welfare-to-Work and other Employment and Training Programs. New York, N.Y.: Manpower Demonstration Research Corporation or Levin; McEwan (2001): Cost-Effectiveness Analysis: Methods and Applications. Thousand Oaks, Calif.: Sage Publications.

[51] Levin; McEwan (2001): Cost-Effectiveness Analysis: Methods and Applications. Thousand Oaks, Calif.: Sage Publications, pp. 46 and references therein.

nancial requirements for an intervention with available resources. This approach is also sometimes referred to as *Cost Feasibility Analysis*.

- *Comparison of actual and budgeted disbursement*, also sometimes referred to as *Cost Variance Analysis*.[52] In an ex-post evaluation, the evaluator can also compare actual expenditures with earlier budget projections. In this way, both the expenditure levels and structure can be compared with projections from the planning stage of an intervention.

In addition to these simple approaches to cost analysis, the analysis can be conducted dynamically by taking time into account (i.e. discounting future costs) as described as part of *financial analysis* in section 4.2.3.

Further generalised, *economic cost analysis* considers costs from a societal perspective as briefly described in section 2.2.4. This most general type of cost analysis is conducted as part of the economic approaches to efficiency analysis described in section 4.3.2 through 4.3.5.

Economic Cost Analysis is based on the concept of *opportunity costs*, i.e. the benefits that cannot be realised because a resource is used for the intervention. This implies the assumption of a reference scenario in which the resource is used otherwise, producing specific benefits.

A common application for opportunity costs is the cost of labour that may be determined by labour-related benefits which a person would enjoy otherwise. Of course, such labour-related benefits should include all effects, positive and negative, which a person enjoys as a consequence of his or her employment. The opportunity cost principle and, more general, the perspective from which costs (and benefits) are assessed sometimes lead to results that are hard to understand by non-economists: relevant costs to society may look very different from the donors' expenditure accounts.

Good cost analysis should be transparent. It is useful to clearly state what type of cost data is presented. Financial accounts usually contain initial budgets, annual budget updates, and actual expenditures at various points in time. In addition, costs can be presented on a cash-out basis or in accounting terms that may contain write-offs over many years. It is important to clearly indicate what type of cost data the analysis is based on, especially if relating cost data from different sources to each other.

In addition, it is useful to check the reliability of the cost data used, for example, by comparing cost data from different sources or overall costs with the sum of costs of an intervention's components. Any mismatch either points to incomplete data in one source or to a misinterpretation by the evaluator.

When assessing aid intervention costs of an aid agency, *a priori*, also institutional overheads of that agency are of interest. They may not be easily attributable to individual interventions in quantitative ways but should nevertheless be addressed transparently.

---

[52] See, e.g., Valadez; Bamberger (1994): Monitoring and Evaluating Social Programs in Developing Countries: a Handbook for Policymakers, Managers, and Researchers. EDI Development Studies. Washington, D.C.: World Bank, pp. 127 and references therein.

2.5. Cost Minimisation and Yield Maximisation

When introducing the concept of efficiency, the related principles of *cost minimisation* and *yield maximisation* are sometimes used.[53] Cost minimisation and yield maximisation represent special cases of efficiency analysis as illustrated in table 2a.

Table 2a. Cost minimisation and yield maximisation as special cases of efficiency analysis

	Alternatives with the same results are compared	*Alternatives with different results are compared*
Alternatives with same costs are compared	**N/A**	**Yield maximisation analysis (a special case of efficiency analysis)**
Alternatives with different costs are compared	**Cost minimisation analysis (a special case of efficiency analysis)**	**General efficiency analysis**

Both principles imply that the respective other side of the equation is held constant:

- *Cost minimisation* searches for the lowest cost alternative that produces the same results, i.e. alternative interventions with the same quantity and quality of results are compared;

- *Yield maximisation* looks for maximum results based on a fixed amount of resources, i.e. alternative interventions with the same costs are compared.

Cost minimisation and yield maximisation have special importance when our grasp of benefits is limited in the following way: if we are only able to rank the benefits produced by different interventions relative to each other but are unable to quantify these benefits, we cannot compare the efficiency of interventions that differ both in benefit and costs. Yet, we can establish an efficiency ranking if costs or benefits of alternatives are held constant.

Imagine, for example, that the results of a number of educational interventions are measured by how much pupils' math test scores increase. We can rank interventions according to how large this test score increase is but we don't know what benefits are produced by different test score increases. Likely, test scores and related benefits are not proportional, i.e. increasing a score from 50 per cent to 51 per cent entails different benefits than increasing the score from 90 per cent to 91 per cent. Therefore, it is difficult to decide whether an intervention producing twice the effect with twice the cost is better or worse.

The only exceptions to this dilemma are the special cases of cost minimisation and yield maximisation:

- *Cost minimisation compares alternatives that produce the same test score increase. Obviously, the least costly option is most efficient: it produces the same benefits with less cost.*

---

[53] See, for example, BMF (2001): Allgemeine Verwaltungsvorschriften (VV-BHO), VV-BHO §7, H 05 01. Bundeshaushaltsordnung (BHO): BMF.

- *Yield maximisation compares alternatives that have the same costs. Obviously, the option that leads to the largest test score increase is most efficient: it produces most benefits with the same cost.[54]*

In technical terms, the optimal interventions identified in this way dominate the other interventions. These, however, remain special cases since usually interventions will differ both in terms of costs and results.

Rather than solving the dilemma described above, cost minimisation and yield maximisation represent exceptional cases in which outcomes can be ranked even if benefits are related to observables in a non-linear way. We will return to these complications and to this example when assessing Cost-Effectiveness Analysis in section 4.3.4.

2.6. Efficiency as Rationale for Decision-Making

Efficiency is a powerful concept. In theory, welfare can be maximised based on efficiency information alone and efficiency would therefore represent the most important criterion in appraisals and evaluations. In practice, however, several factors tend to reduce the potential of efficiency analyses.

If total welfare gains – or net benefits – associated with two or more alternative interventions can be determined, efficiency analysis has great power as a criterion for informed decision-making.

First, if resources are unconstrained, all interventions with positive net benefits should be implemented.[55] Second, if budget, capacity, political or other constraints exist that make a further selection of interventions necessary, the *fundamental principle*[56] of Cost-Benefit Analysis applies:

In any choice situation, select the (policy) alternative that produces the greatest net benefit.

Thus, if there are more interventions (with net benefit larger than zero) than can be implemented, the suggested rational choice is to select the combination of interventions that maximises the total net benefit. In this way, the largest possible welfare improvements are generated. All other combinations of interventions would lead to less welfare improvement and are therefore discarded.

Only the concept of efficiency can support decision-making in this way.

For example, effectiveness and impact are often assessed relative to the stated objectives of an intervention. This *relative* measure of effectiveness depends on how ambitious or convenient effectiveness targets are. Decision-making based on relative effectiveness would be unfair in the sense that it would favour interventions with less ambitious targets.

Even if effectiveness (and impact) are assessed in absolute terms (not relative to targets), effectiveness analysis results alone do not yield sufficient information for welfare-optimising

---

[54] For this statement, we have made the additional common-sense assumption that benefits grow monotonously with increasing test scores.

[55] Actually determination of net benefits is based on direct and indirect assumptions about alternatives (i.e., when assessing opportunity costs and choosing a discount rate). Depending on these assumptions, the net benefit for a given intervention may be positive or negative.

[56] Stokey; Zeckhauser (1978): A Primer for Policy Analysis. New York: W. W. Norton, p. 137, cited by: Gramlich (1990): A Guide to Benefit-Cost Analysis. Englewood Cliffs, N.J.: Prentice-Hall, p. 33.

decision-making. For example, if project A saved ten lives and project B saved 50 lives, we have crucial information in our hands but still don't know which project we should scale up to optimise welfare.[57] In order to take that decision, we need additional information about costs. Note that this expands the effectiveness analysis into an assessment of efficiency. Based on transformation efficiency considerations, we can now conclude that if project A costs less than 20 per cent of project B, scaling up A would be better. Otherwise, scaling up B would be preferable. Without any information about costs, i.e. based on effectiveness analysis results alone, the comparison between the two projects is not meaningful.

Relevance, another evaluation criterion, is an important indicator for how successful an aid intervention is likely to be. Interventions responding to existing needs and designed in line with policies and priorities of target groups, recipients and donors are more likely to be successful than others. Nevertheless, it is easy to imagine an example of low relevance and high efficiency, for example if a new product is introduced that donors, recipients and beneficiaries are still unaware of or if priorities and policies do not adequately reflect needs.

Similar to the argument made above, high sustainability is an important indicator but not a necessary condition for high efficiency. If the benefits of an intervention continue after the intervention ends, additional welfare continues to accumulate over time. This is likely to lead to large effects and high efficiency. However, information about sustainability alone cannot replace efficiency analysis since sustainable benefits may be more than offset by the magnitude of one-time or short-term benefits. Consider the following counterexample: a sustainable intervention with an upfront investment of 1,000 euros leads to net benefits of 100 euros per year ever after. Consider, compared to that, the unsustainable option of short-term investing the same amount in a way that leads to an immediate but one-time net benefit of 1500 euros. For a discount rate of ten per cent, the discounted net benefit is zero in the sustainable option and +500 euros in the unsustainable case, making the latter option more suitable for optimising welfare, i.e. more efficient.

In summary, welfare-optimising decision-making is based on the efficiency criterion. Other evaluation criteria may be able to indicate a *likelihood* for high efficiency but do not guarantee it. If we reliably knew the efficiency of an aid intervention, that information would suffice for welfare-optimising decision-making and the other evaluation criteria would not be needed, i.e. the efficiency criterion would override all other evaluation criteria.

This ideal situation is, however, very unrealistic. Usually, efficiency analysis is based on simplifications and approximations that decrease the reliability of analysis results. In addition, only a certain type of efficiency analysis has the necessary scope to reflect the benefits produced with entire aid interventions. We will refer to this type of analysis as *level 2 analysis* when introducing efficiency analysis levels in section 3.1.1.

Therefore, in practice, efficiency analysis usually does not dictate decisions. Other evaluation criteria are required to complement or substitute for lack of reliability or lack of scope of efficiency analysis.

In spite of these limitations in practical application, the *concept* of efficiency remains of crucial importance for rational decision-making and for informing a welfare-maximising strategy. Reflecting our own observations, the frequency with which aid programming is driven by criteria unrelated to efficiency, for example by political factors, is likely to explain some of the problem of ineffective aid.

---

[57] For our simple considerations, we assume that both projects are scalable at constant transformation efficiency.

2.7. Analysis Perspectives

Appraisal and evaluation questions can generally be asked from different perspectives and the analysis result can depend, sometimes drastically, on the perspective chosen.

For example, when assessing relevance, an intervention can be highly relevant from the donor perspective since perfectly in line with donor policy but, at the same time, lack relevance from the perspective of the partner country that might have other development priorities.[58]

Similarly, effects sizes and costs depend on the analysis perspective chosen.

Choosing a perspective can be considered equivalent to drawing a boundary around a selected part of a society and thereby defining a specific socio-economic system. All changes within that system are assessed in detail while changes outside the system are only considered by what crosses the system's boundaries.

The costs of an intervention are the net inputs provided by all individuals and institutions in the system (plus inputs crossing the borders) and the effects are the attributed net changes on all individuals and institutions in the system (again including border effects).

Keeping this in mind, it is easy to understand that different choices of perspective lead to different costs and effect sizes and, thereby, also to different efficiencies.

A typical analysis perspective used in investment analysis is that of a private-sector entity, for example a company created for implementing a business idea. We will refer to this analysis perspective as the business perspective.

In this simple case, all cash outflows, for example initial investments in equipment and goods or operating costs such as salaries, rent, maintenance and repayments and interest on loans are considered costs. All cash inflows, such as sales revenues and subsidies received are considered benefits. Efficiency analysis from the business perspective is based on the analysis of discounted net cash flows and the resulting expressions that describe the internal efficiency of the business idea under scrutiny are, among other, the net present value (NPV) and the internal rate of return (IRR). This analysis perspective is ideal for investors interested in the monetary return on their investment.

Usually, the business perspective does not yield sufficient information for comprehensively assessing the development efficiency of aid interventions. It does not take into account the effects a business idea has on society, for example the net effects on people employed, net environmental effects and net financial effects from a societal perspective.

In order to be able to include these effects, the analysis must be conducted from a perspective that encompasses all and everything that, directly or indirectly, is involved or affected by the intervention, for example the perspective of the entire society in which the aid intervention takes place.

Such a welfare perspective includes all costs and benefits to society as a whole. If analysed from this perspective, costs and benefits are defined differently than from a company's perspective. For example, salaries paid to staff to implement a business idea are considered costs from a company's perspective. From the perspective of the society, jobs created are assessed by their cost and benefit contribution to society.

Many consequences of aid interventions that are not considered from a business perspective are now considered. These include *external* effects, i.e. consequences that are experienced

---

[58] When assessing relevance, both perspectives (and possibly others) are usually used.

by society but not by a company, and *indirect* effects, i.e. effects, large or small, that happen as a societal or economic reaction to the direct results of an intervention. External effects can be positive or negative, such as improvements or damages to the environment. Examples for indirect effect are increased competition, reduced brain drain and product price changes caused by an intervention.

Other consequences may be relevant only from the individual perspective. These include *transfers* that affect a company but are balanced out within the larger, macroeconomic perspective. If, for example, tax burden is shifted from the company under consideration to another, this represents a benefit to that company but not to society.

Apart from these perspectives, many other analysis perspectives are possible as well. In donor-driven aid interventions, the donor country's society or, in a reduced form, the donor's implementing agency together with the partner country's society could be included into the analysis. In this way, the operational efficiency of the donor agency and its strategic efficiency in selecting the most promising interventions would be integrated with the efficiency of its aid interventions.

Simpler analysis perspectives may also be useful. Choosing the perspective of the ultimate beneficiaries of an aid intervention may reveal whether their participation in an intervention is attractive from their point of view. If, for example, the perceived value of invested time is greater than the perceived benefits, beneficiaries' participation is uncertain unless the intervention design is adapted.

Often, utilising a single analysis perspective will not provide a sufficient picture and several analysis perspectives must be used in parallel. An example is Cost-Benefit Analysis that, if applied only from a society-wide perspective, will hide transfers of welfare between groups within that society. These transfers can be made visible by conducting additional Cost-Benefit Analyses for those groups.

While some of the analysis methods presented later in this report are intended for certain specific perspectives, we do not, *a priori*, pre-select any analysis perspective. This section should however have clarified that a well-defined analysis perspective must be chosen before any analysis of efficiency can be conducted in an unambiguous and meaningful way.

3. Assessment Criteria for Efficiency Assessment Methods

A total of 15 distinct methods and approaches for assessing the efficiency of aid interventions have been identified in the course of this study. These are described and assessed in the next chapter.

In order to increase the transparency of our assessment, we have developed a number of characterisation questions that can be grouped into two categories:

- How powerful is a method, i.e. how useful and reliable are analysis results?
- What are data, resource and skill requirements?

Within each category, we have developed one or more questions that are presented below. In order to also allow for a quick and rough overview, we include simple multiple-choice answer options for each question.

Two points are important to keep in mind in view of assessments within these categories.

- These assessments reflect not only the feedback received from interviewees and the views expressed in books and articles we have reviewed but also our own reasoning and judgement.

- Some methods can be applied with several degrees of rigour or in different ways effectively changing our assessment. In such cases, we chose the assessment that fits most applications and have added a remark in the accompanying text.

Furthermore, we have deliberately decided to not include two criteria that describe the applicability of the methods presented because of the following reasons.

- *Timing of analysis.* All methods described in this report can be applied before, during, or after an intervention takes place. Obviously, ex-ante appraisals cannot be based on empirical data of that intervention but need to extrapolate and plan ahead. The same is true for evaluations during interventions, although some empirical data can be added. Evaluations conducted at the end of interventions or ex-post evaluations conducted several years after that are mostly based on empirical data but may still contain an extrapolation element if long-term effects are of importance.

  Since the answer to the question "At what point in time relative to an intervention is the method applicable?" would yield very similar results for all interventions described in this report, we have decided to not include this characterisation criterion.

- *Classification of aid interventions.* At the design stage of this study, we planned to classify aid interventions based on the applicability of different efficiency analysis methods.[59] Such a classification would obviously be useful for practitioners who could use it to decide which of the methods of this report to consider for a given intervention. Our initial idea was motivated by the observation that certain methods are applied more often to some thematic sectors or specific aid modalities rather than to others.

  A classification scheme that is both informative and useful to practitioners seems however difficult to construct. Schemes suggested in expert interviews and identified in the literature roughly fall into two classes:

  - *Data requirement schemes.* Several classification schemes are simply based on what type of data is needed for the analysis. Most of these differentiate between qualitative and quantitative data and sometimes further distinguish quantitative data into non-monetary and monetary data.[60]

    However, these schemes do not solve our problem since they are simply another way of saying "you can apply this method when this method can be applied, i.e. when all data needed is available." While we follow a very similar approach in this report, we do not refer to it as a classification scheme.

---

[59] See Palenberg (2009): Tools and Methods for Evaluating Aid Efficiency: Inception Report. Bonn: BMZ, pp. 10.

[60] For example, see European Commission (1997): ECO-Fin Manual. Brussels: European Commission, p. 10. Downloaded from http://ec.europa.eu/europeaid/multimedia/publications/publications/manuals-tools/t103_en.htm last visited on 28 May 2010.

- *Impractical schemes.* Other schemes introduce new terminology and suggest classifications that do not relate to any existing thematic, functional or other classification but represent new classifications of aid interventions.[61]

  Since practitioners do not know *a priori* where their intervention belongs in a new classification scheme, the problem is just shifted and not solved. Instead, now, analysis is needed to tell practitioners what efficiency analysis class their intervention is in.

In addition, simply mapping efficiency assessment methods to existing classification schemes is overly simplistic. For example, infrastructure projects are often considered suitable candidates for Cost-Benefit Analysis, while sectors such as education are not. If, however, the main consequences of infrastructure projects are social and environmental in nature, Cost-Benefit Analysis may require more sophisticated valuation approaches. On the other hand, if educational programmes have undergone scientific impact assessment, conducting Cost-Effectiveness Analysis is comparatively easy.

Similarly, the applicability of efficiency analysis methods to different aid modalities cannot be easily mapped. Most methods described in this report are usually applied on the project level. In most cases, however, application to more aggregated aid modalities, such as bilateral programmes and even programme-based approaches, policy loans, budget support, and sector-wide approaches, may be possible as well.

Based on these observations, we have decided to proceed without a classification scheme.[62]

Instead, we characterise efficiency assessment methods by their analytic power and by their requirements. Based on the listed data, time and skill requirements, practitioners can determine which assessment methods can be applied to a specific intervention. In this way, each method is characterised by what it requires, not only on the data level, but also regarding work time and skills.

3.1. Analytic Power of Efficiency Analysis

We characterise the analytic power of efficiency analysis by four criteria, the most important of which is the level of an efficiency analysis. We introduce analysis levels in the next section and summarise all criteria for characterising analytic power after that.

3.1.1. Levels of Efficiency Analysis

As described in section 2.6, efficiency analysis can be used as a basis for informed decision-making. This, however, requires that the efficiency of an entire aid intervention is assessed

---

[61] See, e.g., GTZ (1993): Anleitung für die ökonomische Beurteilung von Projekten der Technischen Entwicklungszusammenarbeit. Frankfurt am Main: GTZ and, for further discussion of this scheme, Voß (1993): Die Wirtschaftlichkeitskontrolle von Entwicklungsprojekten. Frankfurt am Main: Johann Wolfgang Goethe Universität Diplom

[62] For the reader interested in classification schemes for efficiency analysis, Gary McMahon has developed a simple scheme applicable to technical cooperation projects that we consider the best trade-off between pragmatism and specificity. See McMahon (1997): Applying Economic Analysis to Technical Assistance Projects. Policy Research Working Paper Washingotn, D.C.: World Bank 1749, pp. 8 and p. 38.

and compared to an alternative. Not all efficiency analysis is capable of (or intended to) providing such a comprehensive assessment.

For example, in the evaluation of an ongoing intervention, decision-makers may be more interested in finding ways to improve efficiency rather than to compare the efficiency of different interventions. Or decision-makers may be more interested in a simple description of an intervention's costs and benefits without further conclusions.

Obviously, a number of quite different evaluation tasks regarding efficiency exist, each requiring a specific type of analysis. For the purposes of this report, the following three basic task descriptions are useful:

- Assessing the efficiency of an aid intervention in a way that it can be compared with alternatives or benchmarks;

- Identifying efficiency improvement potential within an aid intervention; or

- Describing and providing an opinion on some efficiency-related aspects of an aid intervention.

Obviously, different analysis approaches are required to fulfil these tasks. In our understanding, these approaches differ in their *analytic power*, i.e. in the positive development impact the analysis result itself can trigger. To reflect this, we introduce three *levels of efficiency analysis*.

Analysis methods that *assess the efficiency of an aid intervention in a way that it can be compared with alternatives or benchmarks* have, in our view, the highest analytic power since they can inform strategic optimisation of aid portfolios. Switching from low to high efficiency aid interventions can produce step changes in overall development efficiency. We therefore refer to this type of analysis as *level 2 efficiency analysis*.

In order to produce a reliable estimate of the efficiency of an entire aid intervention in a comparable way, the analysis must satisfy the following criteria:

- The analysis *must cover and provide a reliable estimate of all major benefits and costs* associated with an intervention. This implies that the analysis cannot ignore any important contribution to benefits or costs in the chosen perspective; and

- The analysis *must compare alternatives* by explicitly assessing the overall allocation efficiency of the intervention at hand and that of other, hypothetical or real, interventions. Alternatively, the analysis can produce a standardised indicator for the efficiency of the intervention at hand that has already been calculated for alternative interventions.

If, instead, the focus lies on *identifying efficiency improvement potential within an aid intervention*, leaner analysis may be sufficient. This analysis does not have to cover entire interventions but can focus on some aspects. We will refer to this type of analysis as *level 1 efficiency analysis*. Typical level 1 analyses look at partial efficiency of implementation processes, costs of inputs, conversion of inputs into outputs or conversion of outputs into outcomes. While a typical level 2 analysis can lead to strategic improvements, level 1 analysis usually identifies operative improvement potential.

To identify efficiency improvement potential, level 1 analysis needs to fulfil the following set of requirements:

- The analysis must provide a reliable *partial assessment of efficiency*. Typical examples are analysing the efficiency of an entire component of an intervention, analysis that focuses on just one cause-effect conversion along an intervention's results chain (e.g. only considering conversion of inputs into outputs) or analysis that excludes important cost or benefit categories;

- The analysis must show that the efficiency-related aspect that is analysed can actually be improved. This can, for example, be achieved by benchmarking this efficiency-related aspect across different interventions;

- The analysis must make plausible that overall efficiency will improve if the analysed aspect is improved, e.g. by rationalising that all other things remain equal (*ceteris paribus*) or by additional consideration of effects not reflected in the partial assessment.

Level 1 analysis has less analytic power than level 2 analysis, since it does not provide an overall estimate of the efficiency of an aid intervention and, therefore, provides little assistance for choosing between interventions.

However, since level 1 analysis usually focuses on a specific detail related to overall efficiency, it may provide more insight into how to address specific, efficiency-related challenges within an aid intervention.

Level 1 and level 2 analyses are not exclusive and most level 2 analysis methods will include some level 1 analysis as well. Both types of analysis serve specific purposes: level 2 analysis can assist in selecting the most efficient aid interventions while level 1 analysis is needed for optimising efficiency within given interventions.

Finally, some efficiency analyses neither qualify for level 2 nor for level 1 analysis.

This type of analysis cannot compare the efficiency of aid interventions with each other, nor can it reliably identify specific efficiency-related improvement potential within interventions. Such assessments can, for example, be personal judgements on efficiency or can focus on only one of costs or benefits. We have described the related evaluation task as *describing and providing an opinion on some efficiency-related aspects of an aid intervention* and refer to this type of analysis as *level 0 analysis* or *purely descriptive observation*.

3.1.2. Criteria for Assessing Analytic Power

Our key criterion for the assessment of analytic power of efficiency analysis methods is the *level of efficiency analysis* described above. Apart from describing this criterion for each method, we will select one of the following three multiple-choice options in each case:

- Level 2 analysis;

- Level 1 analysis; or

- Descriptive or level 0 analysis.

In addition to this criterion, we describe the analytic power of efficiency analysis methods by three additional criteria.

As a second criterion, we assess the *degree to which methods are clearly and unambiguously defined*. This criterion examines the degree of available guidance for a specific efficiency assessment method, as well as how much this guidance varies from source to source.

Such guidance may, for example, be provided in academic literature, textbooks or evaluation manuals of aid agencies. In addition to a more detailed description, we select one of the following options for this criterion for each method that is assessed:

- Clear and established analysis procedures exist;

- Some guidance exists; or

- There is little or no established guidance.

A third criterion is the *level of evaluator subjectivity* of analysis results. Assessments of efficiency often contain personal judgements by the evaluator. These judgements can be direct, e.g. when an evaluator estimates effect sizes, or they can enter the analysis by means of professional assumptions, e.g. through the evaluator's choice of calculation or approximation methods. When assessing this criterion we express our estimate of how much we expect analysis results to vary if different evaluators conducted the same analysis under *ceteris paribus* conditions.

In addition to our description, we select one of the following options for this criterion:

- Results obtained by different evaluators are expected to be very similar;

- Results obtained by different evaluators may vary somewhat; or

- Different evaluators can come to entirely different conclusions.

As a fourth criterion for analytic power, we describe the *degree to which a method is participatory* with respect to stakeholders.

In this report, we use the term *stakeholder* generically for the large and diverse group of donors and their agencies, intermediary implementers, and all organisations and individuals having an interest in or experiencing effects of an intervention, the latter obviously including intermediary and ultimate beneficiaries. When using the term stakeholder, it may refer to some or all of the above.

With the *participation* criterion, we assess whether stakeholders are involved at all and the quality of the involvement. Some efficiency analysis methods do not require any stakeholder involvement. Other methods involve stakeholders when collecting data in predefined formats, i.e. when estimating effect sizes by recall methods or by stakeholder estimates. Still other methods are fundamentally based on what criteria are considered important by stakeholders and how these criteria are valued by them. In addition to a more detailed description for each method, we provide a simple overview of our assessment by means of the following three options:

- Stakeholders participate both in defining analysis criteria and in their valuation;

- Stakeholder input is restricted to data gathering along established analysis criteria; or

- The analysis can be conducted without any significant stakeholder involvement.

For the first three assessment criteria, our notion for the magnitude of analytic power decreases in the order of the options that are presented. We consider level 2 analysis to be more powerful than level 1 analysis which, in turn, seems more powerful to us than a descriptive assessment. Similarly, we find clear and unambiguous guidance preferable to some or no guidance. We also feel that methods that produce different results if conducted by different evaluators are less useful than more objective methods.

In contrast to these valuations, we do not connect a value statement with the fourth criterion of stakeholder participation. In our opinion, the degree of stakeholder participation may be of great importance but does not automatically contribute to higher evaluation quality. We merely describe this criterion and leave the value statement to the reader.

3.2. Criteria for Data Requirements

With *data requirements*, we refer to the data the analysis is based on. This may be data available to the evaluator before the evaluation begins, or it may be intermediate data that is generated during the analysis process.

We differentiate two fundamental data types: qualitative and quantitative data.

- With *qualitative data* we refer to all types of information that is described by words and not by numbers.

- *Quantitative data* is all data that is described by numbers.[63] Quantitative data can either be generated by simple counting, e.g. the number of units produced, or by conversion of qualitative data, e.g. through quantification and normalisation of qualitative ratings.[64]

  A special case of quantitative data is *financial data* as found in financial statements and reports. Complementarily, we use the term *numerical data* to describe quantitative but non-financial data.

Several methods for assessing efficiency rely on costs and benefits that are expressed by monetary data, i.e. quantitative data counted in cash units. For doing this, these methods express costs and benefits originally described by any of the above data types by monetary data. This process is called *monetarisation* and, obviously, implies some assumptions that will be discussed later in this report. In order to characterise the data needs of such methods as well, we introduce a third data type which we refer to as *monetarisable data*.

- *Monetarisable data* is qualitative or quantitative data that can be expressed in monetary units based on reasonable assumptions.

As mentioned earlier, monetarisation is a standard procedure but can be tricky if opinions about reasonable assumptions vary. This implies that, for these cases, the assessment of monetarisability depends on the person asked as well as on the intervention itself.

While the three basic data categories described earlier (qualitative, numerical and financial) are mutually exclusive and cover all forms of data, the criterion of data monetarisability represents a different categorisation[65] that we use if methods require such data input.

---

[63] A refined but more complicated definition might differentiate between the scales used for describing quantitative and qualitative data. *Nominal* and *ordinal* measures, even if expressed in numbers, could still be considered qualitative data since no meaningful calculus is possible. Only measures defined on a *ratio* scale where the distance between numbers and the zero-point of the scale have a well-defined meaning, would then be considered quantitative data.

[64] For a description of several methods to generate ratio scales from ordinal scales see, for example, Yoon; Hwang (1995): Multiple Attribute Decision Making : an Introduction. Sage University Papers Series. Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage Publications 07, pp.14 or Hoffmeister (2008): Investitionsrechnung und Nutzwertanalyse. Berlin: Berliner Wissenschafts-Verlag, pp. 289.

[65] That categorisation has only two categories: monetarisable and non-monetarisable. We use only the positive criterion.

Most analysis methods based on qualitative data can also use other data types. Similarly, methods that require numerical data will also accept financial data. When categorising data requirements, we therefore always record *minimum data requirements* along the hierarchy qualitative, numerical and financial rather than listing all possible data types.

In addition to data types, we also describe what level of an intervention's results chain the data originates from.

We differentiate between data on the input, output and outcome-levels. Since it is sometimes difficult to discern outcome and impact level data, we simply include impact-level data into the outcome-level data category.

In summary, this leads to the data requirement matrix shown in table 3a.

Table 3a. Data requirement matrix

		Data type			
		Qualitative	Numerical	Financial	Monetarisable
Data level	Input-level (minimum requirement)				
	Output-level (minimum requirement)				
	Outcome-level (minimum requirement)				

3.3. Criteria for Resource Requirements

Resource requirements are roughly estimated by the *work time* needed for conducting the analysis.

With *work time* we refer to the actual analysis time spent by the evaluator on the assignment. We do not consider the total time needed until the analysis is completed because that may include periods of inactivity.

We have chosen time as the unit of analysis rather than cost since the latter would imply additional assumptions regarding per diem rates.

Nevertheless, the assessment of time needed remains our shakiest assessment criterion. In addition to our own inaccuracies in judgement, the time needed to apply a method is influenced by factors such as the type of intervention that is assessed and the analysis depth and detail. We have therefore created the following very rough multiple choice options and will frequently use work time ranges based on these:

- About a day or less;
- Several days;
- Several weeks;
- Many weeks.

In addition, we also assess the time requirements for stakeholders that represent an opportunity cost for them. Here, the options are:

- None (if the analysis can be conducted without any significant stakeholder involvement);
- A few hours or less per stakeholder involved in the assessment;
- About a day per stakeholder involved in the assessment;

- More than a day per stakeholder involved in the assessment.

Apart from time investments, we do not assess any other resource requirement for the analysis methods presented.

3.4. Criteria for Skill Requirements

Additional special skills may be needed to use some of the methods described in this report.

When characterising skills, a basic evaluation skill set is always needed. Such basic skills are, for example, the ability to reason logically and to conduct basic analysis, information gathering by interviewing and other qualitative methods and report writing and presentation of results.

When describing skill needs, we only refer to special skills that exceed this basic package. The special skills needed for conducting the different types of efficiency analyses presented in this report are:

- Theory-based evaluation skills;

- Deep thematic expertise and long-term evaluation experience;

- Strong qualitative and mixed-method analysis and moderating skills;

- Skills in basic economic and financial analysis;

- Advanced skills in economic analysis;

- Advanced statistical analysis skills; and

- Thorough knowledge of decision theory and utility theory.

For most methods, and unless specified otherwise, both the evaluator conducting the analysis and those commissioning the evaluation require these skills.

4. Characterisation of Methods and Tools for Assessing Aid Efficiency

In what follows, we present methods for assessing the efficiency of aid interventions along the criteria described in the previous chapter. Some methods have similarities to others and sometimes we have chosen to describe these methods in parallel in one section. Overall, we have identified 15 distinct methods that we describe in the following 14 sections.

Our understanding of what constitutes a method is rather flexible. While some of the approaches listed are fully developed methods, others can be better characterised as tools and still others simply as process descriptions. Nevertheless, we refer to all of them as methods.

We present the methods in the order of increasing analytic power, starting with those that are entirely descriptive (level 0), followed by those that identify efficiency improvement potentials (level 1) and those that are capable of comparing entire aid interventions in terms of their relative efficiency (level 2). Within these groups we have roughly ordered the methods in a way that we felt made subsequent explanation easier.

For every method, we begin with a very brief synthesis and a tabular summary of our assessment. This is followed by a more detailed description of the method, our assessment of

its analytic power and of the analysis requirements. In case we have identified good practices or extensions to the method's basic design we present those as well. We conclude each assessment with suggestions for further reading.

4.1. Entirely Descriptive Methods (Level 0 Analysis)

All approaches on this level are entirely descriptive. As such, they don't reveal much about how an intervention works, how its efficiency can be improved, and whether its efficiency is higher or lower than that of alternatives.

Overall, we have identified two distinctively different methods that are described in the following sections:

- Expert Judgement (section 4.1.1) and
- Specific Evaluation Questions on Efficiency (section 4.1.2).

4.1.1. Expert Judgement

An expert judgement on efficiency is the expressed professional opinion of an evaluator on efficiency-related aspects. Although frequently applied, expert judgements depend on the evaluators' own reference frames and different evaluators may come up with different expert judgements for the same intervention. This built-in subjectivity and the fact that expert judgements often lack supporting data or rationale give them only limited analytic power. Our assessment is summarised in table 4a and is explained in more detail in the subsequent text.

Table 4a. Analytic power and analysis requirements for expert judgement

Expert Judgement: Analytic Power				
What is the method's analysis level?	Level 0 (describes and provides an opinion on some efficiency-related aspects of an aid intervention)			
To what degree is the method clearly and unambiguously defined?	There is little or no established guidance			
How independent is the analysis result from the choice of a specific evaluator under *ceteris paribus* conditions?	Different evaluators can come to entirely different conclusions			
How participatory is the method?	The analysis can be conducted without any significant stakeholder involvement			
Expert Judgement: Analysis Requirements				
		Qualitative	Numerical	Financial	Monetarisable
Data	Input-level (minimum requirement)	X			
	Output-level (minimum requirement)	X			
	Outcome-level (minimum requirement)	X			
Time	Overall analysis time needed for evaluator	About a day or less			
	Overall time requirements for stakeholders	None (The analysis can be conducted without any significant stakeholder involvement)			
Skills	Special skills needed	Deep thematic expertise and long-term evaluation experience			

Description. An evaluator expresses his or her professional opinion on the efficiency of a part or of an entire aid intervention. Expert judgements often result in general value statements such as, for example, good, satisfactory or unsatisfactory efficiency.

While the rationale for arriving at an expert judgement on efficiency is sometimes provided when the judgement is presented, more often empirical evidence or explicit logical reasoning is lacking.

Since expert judgements are understood in several different ways, it is important to clarify what is specifically meant here.

- Many evaluation methods involve some degree of expert judgement. Here, we restrict ourselves to the case that the underlying analysis does not involve any other method described in this report.

- Furthermore, the type of expert judgement described in this section is conducted by the evaluator himself. This delineates this approach from expert interviews and expert peer reviews that aim at gathering the feedback of experts other than the evaluator.

Instead, this approach builds on the evaluator's own and personal reference frame and on his or her experience and expectations regarding the intervention under assessment.

A related approach is that of *connoisseurship evaluation* introduced by Elliot Eisner in 1985.[66] This approach places special emphasis on the sensitivity of a connoisseur in discerning subtleties that others may not witness and is similar to the methods used by professional wine tasters or literature and art critics.[67]

A report by Sida[68] on the quality of its own evaluations provides a typical example of an expert judgement statement, taken from a programme evaluation report:

As to the analysis of cost in relation to outputs and outcomes as revealed by the accounts for the Swedish contribution and the detailed scrutiny of each project, the results yielded must on the whole be said to give good value for money.

The Sida study authors comment on this:

As the evidence behind this statement is not given in the report, the reader cannot assess the validity of the assessment. The case is not unique. All too often, conclusions like the one above are presented without supporting data.

One reason for this is the very way in which the assessment is performed, leaving the choice of how the judgement is derived entirely up to the evaluator. This is compounded by the fact that none or very little guidance is available to evaluators for performing expert judgements. In many cases reviewed, no guidance other than reference to general evaluation principles was provided. Harvey Averch describes this lack of guidance in a chapter on using expert judgement in programme evaluation as follows:

---

[66] Eisner (1985): The Art of Educational Evaluation : a Personal View. London; Philadelphia: Falmer Press

[67] For a description of this approach see, for example, Patton (2002): Qualitative Research and Evaluation Methods. Thousand Oaks, Calif.: Sage Publications, pp. 172; Stufflebeam; Shinkfield (2007): Evaluation Theory, Models, and Applications. San Francisco: Jossey-Bass, pp. 184

[68] Forss; Vedung; Kruse; Mwaiselage; Nilsdotter (2008): Are Sida Evaluations Good Enough? An Assessment of 34 Evaluation Reports. Sida Studies in Evaluation. Stockholm: Swedish International Development Cooperation Agency, p. 33

Procedures [to be used for expert judgement] unique to evaluation purposes have not yet been developed. Rather, expert evaluation procedures must be carefully adapted from standard uses such as parameter estimation and forecasting.[69]

Outside the field of development evaluation, criteria for good expert judgements have been suggested. One set of criteria is, for example,[70]

- *Coherence*, i.e. the judgement cannot contradict itself;

- *Calibration*, i.e. the judgement predicts events with the right statistical probability;

- *Resolution*, i.e. the judgement is as unambiguous as possible; and

- *Reliability*, i.e. another judgement by the same expert would come to the same predictions if based on the same facts.

In order to provide sound expert judgements on the efficiency of aid interventions, evaluators must have the necessary skills and experiences to produce judgements that fulfil these or similar criteria. In addition, evaluators must have access to and absorb sufficient relevant information. If, for example, the outcome-level efficiency of an entire aid intervention is assessed, information about outcomes and inputs is needed.

Considering that, in many instances, evaluators are not specifically chosen with respect to their expertise and track record in judging the efficiency of interventions, some scepticism regarding the reliability of expert judgements statements that lack supporting rationale and do not cite relevant facts may be in place. This scepticism was shared by many evaluation professionals consulted for this study.

In our own review of evaluation reports, we have found many expert judgement statements that are introduced by strong disclaimers regarding the difficulty or impossibility of assessing efficiency for the intervention at hand. While these disclaimers seem warranted, somewhat surprisingly, the judgements claimed difficult were then nevertheless provided. In such cases the reader may have doubts regarding the reliability of that assessment.

In other cases, there was a mismatch between the scope of supporting rationale or data and the scope of the judgement, for example if some aspects of operational efficiency are presented but a judgement regarding efficiency on the outcome-level is made.

To us, the most credible expert judgements were those that were limited in scope or that were built on compelling logical arguments. An example for a judgement limited in scope could be the statement that the use of local consultants would be more efficient in a specific project setting, supported by an analysis of the quality and quantity of work done and the costs related to the use of both types of consultants. A compelling logical argument could be the qualitative comparison of two different aid delivery channels with specific discussion of underlying assumptions made in each case. Such a qualitative discussion based on a rigid analytical framework forms the basis of qualitative programme-theory driven evaluations and of qualitative economic assessment. The author of a World Bank study that investigates the potential for applying economical analysis to technical assistance projects summarises this as follows.

---

[69] Averch (2004): Using Expert Judgment. Handbook of Practical Program Evaluation. Wholey, Hatry and Newcomer. San Francisco: Jossey-Bass, pp. 292

[70] Chan (1982): Expert Judgments under Uncertainty: Some Evidence and Suggestions. Social Science Quarterly 63, pp. 428 explanations are provided by the author of this study.

*Finally, in some technical assistance projects it is genuinely difficult to measure outputs either due to the nature of the assistance or the fact that it is part of a broader number of difficult to disentangle measures. However, it will almost always be possible to conceptualize the problem in economic terms. In this manner, the objectives of the project are at least put in a consistent framework. Accordingly, every project should include **qualitative economic analysis** [italics in original] consisting of the conceptualization of the methods and objectives in order to verify their internal consistency.[...][71]*

Analytic power. We consider expert judgements to produce only descriptive results. As explained earlier and summarised below, expert judgements cannot be compared. Usually, expert judgements also do not identify efficiency improvement potential but rather state an opinion on efficiency.

In addition, as described earlier, little guidance is available to aid evaluators for conducting expert judgements.

An important feature of expert judgements is their individual nature. Since efficiency is judged against the evaluators' own standards and expectations, different evaluators may produce quite different judgements, even if each expert judgement itself is done in a professional and diligent manner. Comparing expert judgements by different evaluators compares individual efficiency standards of these evaluators as much as the judged performance of the interventions against these standards.

The only potential exception from this conclusion is a series of judgements conducted by the same expert. Within this set of interventions, the expert judgements would have some degree of comparability since performed by the same individual and based on the same internal reference system. Nevertheless, efficiency rankings produced by such series of expert judgements would, again, depend on the evaluator.

Since expert judgements are individual in nature, the opinions and judgements of stakeholders enter only indirectly, i.e. through modification of the evaluator's own judgement and may be excluded altogether if the evaluator chooses so.

Analysis requirements. Expert judgements have no clearly defined data requirements. In many cases reviewed for this report, it remained unclear what data was used in what way as a basis for the judgement. We consider qualitative data on the input, output and outcome-levels as the minimum basis for producing a reasonably informed expert judgement.

As remarked by a reviewer of an earlier version of this report, it might be useful to further investigate how expert judgements differ if essentially based on output versus outcome data.

The amount of work time needed for conducting an efficiency expert judgement is difficult to isolate from time requirements for the assessment of other evaluation criteria. In some cases, our impression was that expert judgements were formulated without any additional analysis, resulting in a short paragraph on efficiency entirely based on information needed for other points in an evaluation report. In this case, the extra time needed is minimal and can probably be counted in hours or less. If, however, some analysis or reflection is conducted specifically for the efficiency expert judgement, more time may be needed. We estimate that the extra time for conducting an expert judgement on efficiency usually is a day or less.

---

[71] McMahon (1997): Applying Economic Analysis to Technical Assistance Projects. Policy Research Working Paper Washingotn, D.C.: World Bank 1749, p. 16.

The quality of an expert judgement depends on the degree of experience and professional diligence of the evaluator undertaking that judgement. Most interviewees agreed that a deep thematic expertise in the field of the intervention at hand and methodological expertise in conducting efficiency analyses (i.e. a sound understanding of the concept of efficiency) are necessary. It also seemed useful if the evaluator had assessed the efficiency of several similar interventions in the past, preferably using more thorough methodology.

Good practice and suggestions for extension of the basic design. Many improvements to the basic design of expert judgements can be found in appraisal and evaluation reports and were suggested by experts interviewed for this study. Since the analysis approach itself is largely undefined, some of the suggestions for methodological improvements would actually transform expert judgement into another analysis method described in this report. Here, we restrict ourselves to listing suggestions that would improve the reliability and credibility of expert judgements without changing the general approach.

- One way to strongly increase the credibility of expert judgements is to explicitly describe the reasoning and rationale behind it. As pointed out above, if the basis for the judgement remains unknown, sceptical readers of evaluation reports may assume that no supporting facts or rationale exist altogether and seriously doubt the judgement. Underlying rationale for an expert judgement may, for example, consist in a critical examination of the intervention's theory of change that is qualitatively compared to that of alternative interventions. If the judgement is based on facts, these facts should be presented together with the reasoning that led from these facts to the judgement.

- Even without supporting rationale, for example in the case of connoisseur judgements, choosing an experienced and highly regarded expert may help. On the one hand, more professionalism in conducting the judgement can be expected. On the other hand, a standardisation effect may kick in. If the expert at hand has already conducted many judgements of similar interventions, these judgements can be compared with each other since they are, at least to some extent, all based on the same individual reference frame.[72]

- Another way to increase the credibility of expert judgements is to reduce the scope of the judgement. If the judgement is, for example, based on information that covers only a fraction of a larger intervention, it should be restricted to the efficiency of that part of the intervention. If extrapolations are made these need to be transparently displayed and explained.

- Finally, expert judgement can be used to identify efficiency improvements, i.e. as level 1 methodology, if components of alternative interventions are compared with each other. This extended approach is also reflected in the comparative rating models described in section 4.2.4.

---

[72] This, of course, does not imply that another expert wouldn't come up with a different order of efficiency ratings for the same set of interventions.

Suggested literature

- Chan (1982): Expert Judgments under Uncertainty: Some Evidence and Suggestions. Social Science Quarterly 63, pp. 428.

- Averch (2004): Using Expert Judgment. Handbook of Practical Program Evaluation. J. S. Wholey, H. P. Hatry and K. E. Newcomer. San Francisco: Jossey-Bass, pp. 292.

- Eisner (2004): The Roots of Connoisseurship and Criticism - A Personal Journey. Evaluation Roots - Tracing Theorists' Views and Influences. M. C. Alkin. Thousands Oaks, CA: Sage Publications, pp. 196.

- Stufflebeam; Shinkfield (2007): Evaluation Theory, Models, and Applications. San Francisco: Jossey-Bass, pp. 184.

4.1.2. Specific Evaluation Questions on Efficiency

This approach is based on a set of specific theory-based evaluation questions that can be answered by the evaluator based on straightforward data gathering. In contrast to expert judgments, this approach measures efficiency-related information in a structured way. Its analytic power is nevertheless low since it does not usually identify any efficiency improvement potentials. An interesting extension of this approach is to apply the same set of specific evaluation questions to several similar aid interventions. Our assessment is summarised in table 4b and explained in more detail in the subsequent text.

Table 4b. Analytic power and analysis requirements for specific evaluation questions on efficiency

Specific Evaluation Questions on Efficiency: Analytic Power	
What is the method's analysis level?	Level 0 (describes and provides an opinion on some efficiency-related aspects of an aid intervention)
To what degree is the method clearly and unambiguously defined?	There is little or no established guidance
How independent is the analysis result from the choice of a specific evaluator under *ceteris paribus* conditions?	Results obtained by different evaluators are expected to be very similar
How participatory is the method?	Stakeholder input is restricted to data gathering along established analysis criteria

Specific Evaluation Questions on Efficiency: Analysis Requirements					
		Qualitative	Numerical	Financial	Monetarisable
Data	Input-level (minimum requirement)	X			
	Output-level (minimum requirement)	X			
	Outcome-level (minimum requirement)	X			
Time	Overall analysis time needed for evaluator	Several days to several weeks (without considering the time to develop evaluation questions)			
	Overall time requirements for stakeholders	None to a few hours or less per stakeholder involved in the assessment			
Skills	Special skills needed	Theory-based evaluation skills (for the development of evaluation questions, not for conducting the evaluation)			

Description. The development of efficiency-related evaluation questions is certainly not a new appraisal or evaluation method. Many evaluation guidance documents, most terms of reference and many evaluation inception reports contain such collections of questions.

In this report, we define the approach *specific evaluation questions on efficiency* more narrowly:

- The evaluation questions are unambiguous and specific to the extent that little or no leeway exists in interpreting them in different ways;

- The evaluation questions can be answered based on straightforward data gathering with standard methods such as surveys, interviews or document review and do not require substantial assumptions or interpretation by the evaluator.

Examples for specific efficiency evaluation question are:

- What percentage of ministerial employees attending the programme felt that the results justified the time spent?

- What was the total cost for the training per participant per day, i.e. the unit cost for a day of training of one person?

- How can adoption cost best be defined and measured and, based on that, which program component achieved the lowest adoption cost?

Specific evaluation questions on efficiency are not to be confounded with existing generic sets of questions such as the example questions provided with the OECD DAC evaluation criterion efficiency.[73] Several other generic compilations on efficiency exist, all of which provide important guidance but cannot replace more precise questions that allow for straightforward data gathering and unambiguous answers.

Specific efficiency-related evaluation questions have the advantage that the analysis is standardised. It is likely to lead to similar results even if implemented by different evaluators and results of different evaluations may later be compared to each other.

In addition, this approach has the advantage that it is also applicable to complex interventions or interventions with hard to quantify outcomes.

While being suggested by several experts, we have not seen this approach applied to the analysis of efficiency in practice. Without specific focus on efficiency, budget support evaluation frameworks[74] or the European Commission's Results-Oriented Monitoring framework[75] may illustrate the general idea behind this approach.

Analytic power. Since, in our definition, this approach is based on evaluation questions that can be answered by straightforward data gathering we do not expect it to identify efficiency improvement potential but rather to provide a snapshot of several efficiency-related indicators. Therefore, we categorise this approach as only yielding descriptive results.

---

[73] OECD DAC (2010): Evaluating Development Co-Operation: Summary of Key Norms and Standards. Paris: OECD, p14. Three generic questions are suggested: Were activities cost-efficient? Were objectives achieved on time? Was the programme or project implemented in the most efficient way compared to alternatives?

[74] Hammond (2006): A Framework for Evaluating Budget Support. Koeberle, Stavreski and Walliser. Washington, D.C.: World Bank pp. 91.

[75] European Commission (2009): Results Oriented Monitoring: ROM Handbook. Brussels: European Commission pp. 70.

However, in a second step, the results obtained by evaluations of different interventions can be used to draw further insights as explained in the extension of design discussed below.

The main advantage of this approach, relative to expert judgement, lies in the fact that evaluation results can be expected to be pretty much independent of the individual evaluator, provided that the evaluation questions are answered with professional rigour. With predesigned specific evaluation questions, the weight of developing and choosing adequate methodology and to interpret results is shifted from the shoulders of the evaluator to those commissioning an evaluation.

While many terms of reference, evaluation guidelines and other publications contain sets of efficiency-related evaluation questions, we have found only little documentation that included the level of specificity discussed here. Overall, the literature on programme-theory driven evaluation[76], logic models[77] or simply on developing sound measures[78] may be of assistance when developing this type of efficiency-related evaluation questions.

The degree to which stakeholders influence the analysis result obviously depends on the nature of the evaluation questions. We expect stakeholders to be mainly involved for data gathering purposes.

Analysis requirements. Data requirements depend on the specific evaluation questions. We consider qualitative information on the input, output and outcome-levels as minimum data requirement. It is important, however, that questions are specific and that it is possible to answer them through straightforward data gathering without much interpretational freedom. Some people interviewed felt that this would be easier if the questions required quantitative (numerical) answers.

The work time needed for applying this approach based on existing evaluation questions strongly depends on how difficult and lengthy data gathering procedures are. Most often, this approach will also be embedded in the evaluation along other criteria than efficiency and draws on the results produced. In its shortest version, the entire evaluation, including the assessment of not only efficiency but of all evaluation criteria required about 2 weeks.[79] Work time needs can therefore vary largely and we estimate them to lie between several days and several weeks.

In addition, the formulation of the evaluation questions themselves may require a considerable investment of time if researched thoroughly. This substantial additional time requirement is not included into our above estimate since it may not be required for every single evaluation, for example if it is possible to formulate a set of specific questions for an entire category of interventions (e.g. group training sessions of a certain duration). Time requirements for the question development will usually need to be borne by those planning evaluations.

Time requirements for stakeholders are expected to be modest, i.e. several hours or less per stakeholder involved in the assessment.

---

[76] See, for example, Rogers (2000): Program Theory in Evaluation: Challenges and Opportunities. San Francisco, Calif.: Jossey-Bass.

[77] See, for example, Frechtling (2007): Logic Modeling Methods in Program Evaluation. San Francisco: Jossey-Bass.

[78] See, for example, Weiss (1998): Evaluation: Methods for Studying Programs and Policies. Upper Saddle River, N.J.: Prentice Hall pp. 114.

[79] Work time average for conducting the European Commission's Results Oriented Monitoring (ROM) assessment. Information obtained during interview.

Skills requirements are twofold. On the side of the evaluator, only standard evaluation skills such as interviewing and basic analytical skills are required.

However, when developing the evaluation questions, special skills are needed. In order to derive meaningful evaluation questions that can be answered by straightforward data gathering, a thorough understanding of the theory of change of the intervention at hand needs to be established. This requires advanced skills in theory-based evaluation.

Good practice and suggestions for extension of the basic design. In order to increase the analytic power of this approach, a standardised set of specific (i.e. not generic) questions can be applied in evaluations of different but similar interventions. The results of these separate evaluations can then be compared and may reveal efficiency improvement potential or even point to which interventions may be most efficient.

Suggested literature

- Frechtling (2007): Logic Modeling Methods in Program Evaluation. San Francisco: Jossey-Bass.

- Rogers (2000): Program Theory in Evaluation: Challenges and Opportunities. San Francisco, Calif.: Jossey-Bass.

- European Commission (2009): Results Oriented Monitoring: ROM Handbook. Brussels: European Commission.

4.2. Methods for Identifying Efficiency Improvement Potential (Level 1 Analysis)

In this section, we list a number of methods that are able to identify efficiency improvement potentials. In contrast to the entirely descriptive methods presented earlier, level 1 methods identify efficiency improvement potential by comparing elements of an intervention with alternatives in explicit or implicit ways. In contrast to level 2 analysis methods presented in the next section of this report, level 1 analysis methods do not look at entire interventions.

Overall, we have identified seven distinct level 1 methods that are described in the following five sections:

- Benchmarking of unit costs and benchmarking of other partial efficiency indicators (section 4.2.1);

- Follow the Money (section 4.2.2);

- Financial analysis (section 4.2.3);

- Two different comparative ratings by stakeholders (section 4.2.4); and

- Stakeholder-driven approaches (section 4.2.5).

4.2.1. Benchmarking of Unit Costs and of Other Partial Efficiency Indicators

Benchmarking of unit costs is a powerful efficiency analysis technique that can identify efficiency improvement potential in aid interventions. It compares cost per output, for example the cost per kilometre of road built, across several interventions. Similar analysis can be conducted with other partial efficiency indicators, for example the administrative costs per beneficiary. When conducting either type of these analyses it is crucial to make plausible that the investigated quantity is positively correlated with overall efficiency. Without this last step,

erroneous conclusions can be drawn. Our assessment is summarised in table 4c and explained in more detail in the subsequent text.

Table 4c. Analytic power and analysis requirements for benchmarking of unit costs and of other partial efficiency indicators

Benchmarking of Unit Costs and of other Partial Efficiency Indicators: Analytic Power	
What is the method's analysis level?	Level 1 (identifies efficiency improvement potential within an aid intervention)
To what degree is the method clearly and unambiguously defined?	Some guidance exists
How independent is the analysis result from the choice of a specific evaluator under *ceteris paribus* conditions?	Results obtained by different evaluators are expected to be very similar
How participatory is the method?	The analysis can be conducted without any significant stakeholder involvement

Benchmarking of Unit Costs and of other Partial Efficiency Indicators: Analysis Requirements					
		Qualitative	Numerical	Financial	Monetarisable
Data	Input-level (minimum requirement)		X (other indicators)	X (unit costs)	
	Output-level (minimum requirement)		X		
	Outcome-level (minimum requirement)				
Time	Overall analysis time needed for evaluator	About a day or less to several days			
	Overall time requirements for stakeholders	None (The analysis can be conducted without any significant stakeholder involvement).			
Skills	Special skills needed	-			

Description. The benchmarking of unit costs is a well-known and widely applied method for partial efficiency analysis. Unit costs are defined as the cost associated with "producing" a unit of output without taking into account initial investments or multi-year effects. Unit costs are benchmarked against identically defined unit costs in other interventions or against experience-based benchmarks.

The benchmarking of unit costs itself is only a partial efficiency analysis since it does not cover all outputs of an intervention and does not include any information on how outputs translate into outcomes (and impacts). Therefore, it cannot provide information on the entire intervention without further assumptions or analyses. Unit cost benchmarking compares production efficiency rather than allocation efficiency.

Very similar to unit costs, other quantitative indicators can be defined that, while not measuring overall efficiency, nevertheless measure one important contribution to the overall efficiency of an aid intervention. For example, some efficiency analysis methods suggested by IFAD[80] are based on the calculation and the benchmarking of partial efficiency indicators such as, for example:

- Loan costs per beneficiary;

- Administrative costs per beneficiary;

---

[80] IFAD (2009): Evaluation Manual: Methodology and Processes. Rome: International Fund for Agricultural Development, Office of Evaluation.

- Time until loan was effective; and

- Extension of original closing date and related additional administrative costs.

However, evaluators need to carefully consider the limitations of analyses based on unit costs and other partial efficiency indicators when drawing insights and developing recommendations. In order to provide information about overall efficiency, unit costs and partial efficiency indicators need to be positively correlated to allocation efficiency in the sense that, if other factors remain unchanged (*ceteris paribus*), an increase in these indicators translates into an increase in overall efficiency. These extra considerations are needed since, usually, any recommendation that leads to a change in the partial efficiency indicator will also affect other parts of the intervention not mapped by the indicator.

> *If, for example, the cost per kilometre of road built, the cost per training participant, or the time delay until a project became effective is higher or larger than for other benchmarks, it cannot be concluded that the intervention's allocation efficiency would increase if these unit costs or time delays were reduced: the positive contribution to increased efficiency might as well be more than offset by adverse effects, e.g. by reduction in road and training quality, or by reduced project planning quality.*

In some cases, the necessary complementary analysis described above is fairly straightforward. If a kilometre of road can be built for less money with the same quality and no other adverse effects, the allocation efficiency of this intervention will clearly increase if the lower-cost-option is chosen. In other cases, for example training, the respective effects triggered by low- or high-cost training may be harder to estimate and, possibly, a conclusive statement on allocation efficiency cannot be made. However, even under these circumstances, partial efficiency indicators are very useful for raising red flags and pointing out areas for additional analysis.

These restrictions in utilising benchmarking results for partial efficiency indicators has led some of the experts interviewed for this study to reject this type of analysis altogether. These experts saw too great a risk of explicit or implicit conclusions based on insufficient facts. Indeed, in our own review of evaluation reports that contained unit cost and other partial efficiency indicator benchmarks, observations regarding overall efficiency of interventions were sometimes based on a fragment of the necessary information only.

Most experts interviewed, however, felt that this type of analysis was useful if interpreted with care as described above.

As with benchmarking in general, a comparison of unit costs (or other partial efficiency indicators) is only meaningful if standardised measurement definitions exist. In the case of unit costs this implies an exact definition of what costs are to be included in the calculation, for example:

- *Average operational cost per unit of output* (e.g. fixed and variable operational costs for a given year divided by the number of units produced); or

- *Marginal operational cost per unit of output* (e.g. increase of variable cost needed to produce an additional unit of output).

However, even for these cases, interpretation and accounting standards may differ between institutions. Therefore, special attention has to be paid to compare apples with apples and not with pears when conducting this type of analysis.

An interesting generalisation of these simple annual unit costs are *dynamic prime costs* as, for example, used by KfW.[81] We consider this type of analysis to be very close to financial analysis which is described later in this chapter. Dynamic prime costs are calculated as the sum of discounted costs over the sum of discounted quantity of output and therefore indicate the average net present unit cost required to recover all net present costs. Dynamic prime costs are a useful basis for developing tariff and pricing models.

Analytic power. The benchmarking of unit costs and of other partial efficiency indicators represents a simple and powerful efficiency analysis tool. If treated with analytic care as described above, this type of analysis can identify efficiency improvement potential. We therefore classify it as level 1 analysis.

In some cases, partial efficiency indicators were calculated with explicit reference to agency standards. In other cases, it remained somewhat unclear what standard had been used. Overall, our impression is that only some guidance exists.

However, if standardised measures for partial efficiency indicators exist, analysis results can be expected to be independent from the choice of evaluator.

For this type of analysis, usually stakeholders are not significantly involved.

Analysis requirements. A key advantage of this efficiency analysis method is that its data, resource and skill requirements are comparatively low. This translates into broad applicability.

For the calculation of unit costs, financial cost and numerical output data are needed. For other partial efficiency indicators, only numerical data may be needed. Usually, this information is easily accessible.

Work time requirements for this method are equally low. If the necessary information is available, indicators can be calculated in a matter of hours. However, in order to establish meaningful benchmarks, more time may be needed, either for verifying that indeed all comparisons have been calculated using the same definitions or by even recalculating indicators in alternative scenarios. Overall, we estimate work time requirements for the evaluator to range between a day and several days. No time from stakeholders is required.

For the application of this method the evaluator does not need special skills.

Good practice and suggestions for extension of the basic design. Unit cost benchmarks or comparisons of other partial efficiency indicators are useful if a number of good practices are observed:

- As already described, unit costs and other partial efficiency indicators need to be correlated to the overall efficiency of an intervention. If such further analysis proves to be difficult, it may be useful to leave additional judgement to the decision-maker and present findings in a different way, for example by explaining that, in order to have equal allocation efficiency, the more expensive training needs to generate three times stronger outcome-level effects than the lower cost training.

- The calculation of unit costs or of other partial efficiency indicators needs to be based on standardised definitions. When selecting similar quantities for benchmarking purposes, it must be ensured that these have been calculated based on exactly the

---

[81] See, e.g., Bentlage; Petry (2008): Einzelwirtschaftliche Investitionsanalyse. Frankfurt am Main: KFW Entwicklungsbank, p. 57.

same definitions. If this is not the case, recalculation or at least an estimate of the error made is recommended.

Suggested literature

- *Costs Analysis Step by Step: A How-to Guide for Planners and Providers of Welfare-to-Work and Other Employment and Training Programs*, D.H. Greenberg, U. Appenzeller, MDRC, 1998, pp. 22.

4.2.2. Expenditure Tracking (Follow the Money)

"Expenditure Tracking" or "Follow the Money" are our own names for a pragmatic approach that traces all expenditures associated with an intervention to their outputs and estimates the likelihood that these outputs produce intended or unintended outcomes. This approach systematically and exhaustively scans an intervention for operational efficiency improvement potential by searching for cost minimisation or yield maximisation potential. Our assessment is summarised in table 4d and explained in more detail in the subsequent text.

Table 4d. Analytic power and analysis requirements for Follow the Money

Follow the Money: Analytic Power				
What is the method's analysis level?	Level 1 (identifies efficiency improvement potential within an aid intervention)			
To what degree is the method clearly and unambiguously defined?	There is little or no established guidance			
How independent is the analysis result from the choice of a specific evaluator under *ceteris paribus* conditions?	Results obtained by different evaluators may vary somewhat			
How participatory is the method?	Stakeholder input is restricted to data gathering along established analysis criteria			
Follow the Money: Analysis Requirements				
		Qualitative	Numerical	Financial	Monetarisable
Data	Input-level (minimum requirement)			X	
	Output-level (minimum requirement)	X			
	Outcome-level (minimum requirement)	X			
Time	Overall analysis time needed for evaluator	Several days to several weeks			
	Overall time requirements for stakeholders	A few hours or less per stakeholder involved in the assessment			
Skills	Special skills needed	-			

Description. The Follow the Money approach is a pragmatic yet exhaustive way of identifying efficiency improvement potential. While, to our knowledge, not described as a separate approach in literature, several experts interviewed for this study described approaches that we try to summarise within this approach. In addition, this approach bears strong resemblance to how private sector management consultancies tackle performance optimisation of companies.

In the Follow the Money approach, all expenditures connected with an intervention are tracked and the resulting outputs are recorded. In a second step, these outputs are then assessed according to their magnitude and quality, as well as their likeliness for triggering intended or unintended outcomes. While "following the money" through the system, the evalua-

tor searches for cost minimisation potential (could the same result be achieved in a less costly way?) and for yield maximisation potential (could more result be achieved based on the same costs?) as described in section 2.5.

While the first step is a straightforward and systematic tracking and mapping exercise, the second step may involve third party input and is based on the evaluator's own judgement. No specific methodology is prescribed for addressing the latter question.

Overall, the Follow the Money approach represents a structured framework for exhaustively and systematically considering an intervention.

One interviewee summarised his version of this approach as follows:

> *In the 'Follow the Money' approach, the evaluator starts as a skeptic. Program funds were spent on interventions, but because there was no random selection of participants, no control population, and no statistically sound way of defining the counterfactual, the evaluator knows in advance that formal attribution of impacts to the program intervention will be impossible. Even in such cases, however, it is reasonable for evaluators to look for evidence of impact, recognizing that any evidence that is found will be necessarily weak rather than compelling. In the course of the evaluation, bit by bit, the evaluator might demonstrate that all expenditures were made following the original logic of the program design, the activities were completed and outputs generated, and that some evidence exists that impacts were achieved.*

Analytic power. The Follow the Money approach represents a typical level 1 analysis, capable of identifying efficiency improvement potential without providing an overall assessment of an intervention's efficiency relative to that of others. It is a pragmatic approach that can be used when other, more refined approaches are not applicable.

The main strength of the Follow the Money approach is its ability to systematically and exhaustively screen an entire intervention for wasted resources or obvious inefficiencies. Since all expenditures are tracked, expenditures that cannot be connected to any activity or output automatically surface. Similarly, outputs that are very unlikely to produce any beneficial outcome can easily be identified, for example in the form of unused equipment or irrelevant training. The Follow the Money approach can also guide the evaluation of other criteria than efficiency in terms of what priority areas to focus on. It produces a comprehensive overview of priority expenditure areas.

Apart from the general idea of tracking all expenditures, little methodological guidance exists in how to assess outputs, i.e. in terms of their likeliness to generate intended outcomes, and on how to relate these results back to specific expenditure categories.

Since this last step is left largely to the ingenuity of the evaluator, evaluation results may differ somewhat between evaluators. Usually, some stakeholder input is needed in the assessment.

Analysis requirements. The Follow the Money approach requires financial input data. Ideally, detailed budget and actual expense databases are used. Feedback is also needed on the output and outcome-level but can be qualitative or any other form of data. We record qualitative data as minimum data requirement on these levels.

Analysis time for the Follow the Money approach scales with the size and the complexity of an intervention, although a lower degree of detail may be chosen for larger interventions.

When based on existing, reliable expenditure data, the entire analysis can be conducted in a matter of days rather than weeks. However, when conducted on a very detailed level or if reliable expenditure data first needs to be assembled as part of the analysis, the work time requirements may be several weeks.

Time requirements for stakeholders are usually moderate, e.g. a few hours or less per stakeholder. Individual stakeholders, such as the manager of the intervention, may be considerably stronger involved.

In terms of skills, the Follow the Money approach requires analytic discipline and basic understanding of financial information, but no special evaluation skills.

Suggested literature. We have not identified literature specific to this approach.

4.2.3. Financial Analysis (Microeconomic Investment Analysis)

Financial analysis is a powerful analysis tool for assessing the financial value of an investment. In an aid context, it is useful if private sector elements are part of an aid intervention. For those elements, financial analysis can assess both financial viability and efficiency improvement potential. Our assessment is summarised in table 4e and explained in more detail in the subsequent text.

Table 4e. Analytic power and analysis requirements for financial analysis

Financial Analysis: Analytic Power				
What is the method's analysis level?	Level 1 (identifies efficiency improvement potential within an aid intervention)			
To what degree is the method clearly and unambiguously defined?	Clear and established analysis procedures exist			
How independent is the analysis result from the choice of a specific evaluator under *ceteris paribus* conditions?	Results obtained by different evaluators may vary somewhat			
How participatory is the method?	Stakeholder input is restricted to data gathering along established analysis criteria			
Financial Analysis: Analysis Requirements				
		Qualitative	Numerical	Financial	Monetarisable
Data	Input-level (minimum requirement)			X	
	Output-level (minimum requirement)				
	Outcome-level (minimum requirement)			X	
Time	Overall analysis time needed for evaluator	Several days to several weeks			
	Overall time requirements for stakeholders	A few hours or less per stakeholder involved in the assessment			
Skills	Special skills needed	Basic economic and financial analysis			

Description. Microeconomic investment analysis or simply, financial analysis, calculates the net present value or the internal rate of return of an investment from a business perspective. It is particularly useful for aid interventions that include a clearly delineated business component, for example a public or private service company.

The analysis is based on annual cash flows, i.e. the revenues and expenditures generated by a business component of an aid intervention. Past and future cash flows are converted to

their present value at some point in time, usually to the present year. A discount rate, essentially incorporating the valuation of other investment alternatives, is used for this analysis. Direct comparisons across alternative interventions can be made as long as the same discount rate is used.[82] Depending on the type of business interventions at hand, cash flows are either truncated after a number of years or are continued forever. Present values are then summed up and the net present value (NPV) is calculated. The NPV represents the total financial value of a business idea. In a private sector context, it is usually used to inform investment decisions: all business ideas with NPV larger than zero can be considered profitable and between several investment options, the option with the largest NPV is preferable.

Alternatively, the internal rate of return (IRR) can be calculated, which is simply the constant discount rate at which the net present value is zero. Since the equation for the IRR is nonlinear, multiple solutions may exist and care must be taken in selecting the economically meaningful solution.[83]

Analytic power. Microeconomic investment analysis is a powerful tool for assessing the profitability of a business idea. It can be considered a simple form of Cost-Benefit Analysis that is restricted to cash in- and outflows from a company's perspective.

For aid interventions containing a business element, microeconomic investment analysis is of great importance. If, for example, an aid intervention consists in setting up infrastructure for the provision of basic services, the financial sustainability of the entity responsible for implementation and service provision is crucial for the success of the entire aid intervention. In this sense, microeconomic investment analysis can reveal efficiency improvement potentials. Its main use, however, lies in providing information for design and planning of business elements in aid interventions, e.g. in determining tariffs or prices that allow for financial sustainability.

We categorise microeconomic investment analysis as level 1 analysis. We do not consider it a level 2 analysis since it usually does not map the entire aid intervention. In other words, while the financial sustainability of an intervention's business element may be a necessary condition for development outcomes, it usually is not a sufficient condition.

For interventions without a central business element, microeconomic investment analysis is too limited and non-financial costs and benefits from other perspectives need to be included. This more general approach will be discussed when assessing the Cost-Benefit Analysis in section 4.3.2.

Microeconomic investment analysis is a thoroughly researched field and detailed and established guidance exists. When conducting the analysis, the evaluator nevertheless has to make a number of specific assumptions that may influence the analysis result.

Stakeholders are usually only involved for data gathering purposes. It should be noted, however, that several reviewers of an early version of this report have suggested to also considering a participatory version of the Follow the Money approach that would involve stakeholders in all stages of the assessment process.

Analysis requirements. For microeconomic investment analysis, financial data is needed on the input-level and some financial information on results is required as well. Since quanti-

---

[82] If interventions differ by their risk, this may not be true and, instead, different discount rates representing different risks need to be used.

[83] Both the NPV rule and the potential issues with using rates of return are discussed in more detail in their generalised form when introducing Cost-Benefit Analysis in section 4.3.2.

ties and prices for products and services are involved, we characterise this as financial data on the outcome-level.

The analysis time depends on the size and complexity of the business element and the degree of reliability and detail the analysis aims for. Overall, we estimate an overall work time for the evaluator between several days and several weeks. Time requirements on stakeholders are minimal, i.e. a few hours or less per stakeholder involved.

In order to be able to conduct microeconomic investment analysis, the evaluator needs to possess basic economic and financial analysis skills.

Good practice and suggestions for extension of the basic design. A wealth of literature and experience exists on microeconomic investment analysis, including many good practices. We restrict ourselves to the following two remarks:

- When conducting microeconomic investment analysis, it is important to clearly communicate assumptions made, for example regarding the time horizon and the discount rate utilised. Otherwise, comparisons (for example of IRRs) may be misleading.

- A helpful aid to structuring and visually representing the contributions to one-time financial cash flows are *ROI trees* (where ROI stands for Return on Investment) that are widely used in management consulting. In ROI trees, the return on investment of a business activity is systematically broken down into its different drivers, i.e. into investment and operational results, the latter into revenues and costs, and so forth. This type of additional analysis allows pinpointing root causes and, thereby, informs recommendations for efficiency improvements.

Suggested literature

- Copeland; Koller; Murrin (2000): Valuation: Measuring and Managing the Value of Companies. Wiley Frontiers in Finance. New York: John Wiley.

- European Commission (2008): Guide to Cost-Benefit Analysis of Investment Projects. Brussels: European Commission, pp. 34.

- Bentlage; Petry (2008): Einzelwirtschaftliche Investitionsanalyse. Frankfurt am Main: KFW Entwicklungsbank.

4.2.4. Comparative Ratings by Stakeholders

Comparative ratings by stakeholders record the feedback of those involved or affected by an aid intervention on the efficiency or effectiveness of that intervention, usually by means of surveys or interviews. Stakeholders are asked to rank different options according to efficiency or effectiveness criteria. If effectiveness criteria are used, the evaluator needs to complete the analysis by comparing these ratings to the respective costs.

These methods have been suggested to us by several experts. We have included them into this catalogue although, somewhat surprisingly, we have not seen many applications to the analysis of efficiency. Our assessment is summarised in table 4f and explained in more detail in the subsequent text.

Table 4f. Analytic power and analysis requirements for comparative ratings by stakeholders

Comparative Ratings by Stakeholders: Analytic Power	
What is the method's analysis level?	Level 1 (identifies efficiency improvement potential within an aid intervention)
To what degree is the method clearly and unambiguously defined?	There is little or no established guidance
How independent is the analysis result from the choice of a specific evaluator under *ceteris paribus* conditions?	Results obtained by different evaluators may vary somewhat (if the same set of rating question is used)
How participatory is the method?	Stakeholder input is restricted to data gathering along established analysis criteria

Comparative Ratings by Stakeholders: Analysis Requirements					
		Qualitative	Numerical	Financial	Monetarisable
Data	Input-level (minimum requirement)	X (for efficiency rating)		X (for effectiveness rating)	
	Output-level (minimum requirement)	X			
	Outcome-level (minimum requirement)	X			
Time	Overall analysis time needed for evaluator	Several days			
	Overall time requirements for stakeholders	A few hours or less per stakeholder involved in the assessment			
Skills	Special skills needed	-			

Description. Two methods that have mainly been suggested in interviews are presented in this section. Both are based on ratings by stakeholders that are collected through interviews or surveys. We have not been able to identify many appraisals or evaluations that employed these tools, so they should be considered suggestions by some of the experts interviewed rather than as established methods.

Before discussing these methods in more details, a disambiguation remark is necessary. Comparative ratings by stakeholders bear some resemblance to methods of Multi-Attribute Decision-Making (MADM) that are sometimes also referred to as scoring models, as well as to Cost-Utility Analysis (CUA). Both MADM approaches and CUA are discussed later in this chapter. With respect to those methods, comparative ratings by stakeholders have important methodological differences: while MADM methodology requires assessments to be conducted by a decision-maker (and not a stakeholder group), Cost-Utility Analysis first quantifies all relevant outcomes and, only subsequently, values them by their respective utilities.

Between each other, the two comparative stakeholder rating methods discussed here differ by whether stakeholders are asked to provide an assessment of efficiency or of effectiveness:

- *Comparative rating of efficiency.* Stakeholders are asked to rate specific aspects of an intervention's efficiency against real or hypothetical alternatives.

- *Comparative rating of effectiveness and cost analysis.* With this approach, stakeholders are asked to rate the effectiveness of different alternatives. The evaluator then assesses the costs of each alternative in order to come to a conclusion about efficiency.

With both approaches, care must be taken that the questions asked can be answered based on what stakeholders know and are able to observe. If, for example, surveyed stakeholders are unaware of the costs related to different alternatives, these costs should be provided as information together with the questions or, alternatively, questioning should be restricted to the assessment of effectiveness as in the second method. In a similar vein, questions asked from a donor perspective, for example referring to programmes or strategies may be entirely incomprehensible for local stakeholders. It is therefore important to adapt the questions to the language and the observational context of the stakeholder groups that are surveyed. We also assume that minimum standards for questionnaire design are observed.[84]

Finally, it should be kept in mind that this type of analysis essentially produces qualitative results. Even if ratings are based on numerical scales, twice the score usually does not equal double efficiency or effectiveness. Therefore, it is usually not correct to calculate cost-effectiveness ratios, even though evaluators might feel tempted to do so in the second method. We discuss the underlying scaling issues in more detail when presenting Cost-Effectiveness Analysis in section 4.3.4. If, instead, a conversion from nominal or ordinal scales into ratio scales has been conducted, this should be made explicit as part of the assumptions made in the analysis.[85]

Analytic power. Both comparative rating methods presented here can provide estimates of efficiency and efficiency improvement potentials that – in contrast to methods based on financial data – include an assessment of perceived costs and benefits from the stakeholder perspective. These are not restricted to monetary or other tangible quantities but include all effects experienced by stakeholders.

If a set of questions on underlying reasons for perceived high or low efficiency is included, concrete leads for improving the efficiency of an intervention can be identified. Therefore, we categorise both types of comparative ratings by stakeholders as level 1 analysis. We do not consider these approaches to be able to compare entire interventions (level 2 analysis) since it is unlikely that stakeholders that can judge effects on the ground also have an overview of the entire intervention.

In terms of established guidance, we have not been able to identify literature that specifically describes this general approach while focusing on efficiency and including a comparison.

The specific implementation design of this approach is usually determined by the evaluator. He decides on the specific questions to be asked, designs the survey or conducts the data gathering interviews and analyses cost in the effectiveness-based variant. We assume that guidance as to what aspect of an intervention's efficiency is to be looked at is given to the evaluator. Based on these observations and assumptions, we estimate that results will depend (only) somewhat on the evaluator chosen.

In both approaches, stakeholders are strongly involved and provide their feedback along predefined dimensions.

---

[84] See, for example, Bamberger; Rugh; Mabry (2006): RealWorld Evaluation: Working Under Budget, Time, Data, and Political Constraints. Thousand Oaks: Sage Publications, pp. 240.

[85] For a description of several methods to generate ratio scales from ordinal scales see, for example, Yoon; Hwang (1995): Multiple Attribute Decision Making : an Introduction. Sage University Papers Series. Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage Publications 07 pp. 14 or Hoffmeister (2008): Investitionsrechnung und Nutzwertanalyse. Berlin: Berliner Wissenschafts-Verlag, pp. 289.

Analysis requirements. In case of comparative ratings of efficiency, i.e. the method in which stakeholders directly rate efficiency, no additional data is required for the evaluator. However, stakeholders are likely to require some information for qualified judgements which, at least, should cover qualitative data on the input and output-level. It may be necessary for the evaluator to make this data available to the stakeholders prior to the assessment.

In case of comparative ratings of effectiveness, the evaluator needs to add his or her own cost analysis in order to assess efficiency. Therefore, financial input data is needed.

An advantage of these analysis methods is their low time requirement. We estimate that both methods can be implemented in a matter of days of work time (not including survey wait times). Time requirements for stakeholders exist but are modest. We estimate that the time to participate in a survey or a structured interview should not exceed a few hours.

No special skills are needed for this analysis. The evaluator must be able to design, conduct and analyse surveys or interviews and to perform basic cost analysis in case of the second method.

Good practice and suggestions for extension of the basic design

It should be noted, that this method's description is based on the assumption that the questions and options in surveys and interviews are pre-fabricated by the evaluator.

An interesting extension of this approach was suggested by several experts providing feedback on a draft version of this report by including the option to actively include a participatory category and question design phase into the evaluation process. This is likely to produce more relevant questions but would require more time from stakeholders and the evaluator.

Suggested literature. No documentation specific to this method has been found. However, some of the literature on the related methods for Multiple-Attribute Decision-Making discussed in section 4.3.1 may be useful.

4.2.5. Stakeholder-Driven Approaches

"Stakeholder-driven approaches" is our own catch-all phrase for the little explored field of highly participative efficiency analysis methods. While not explicitly described in textbooks, there is no principal obstacle for involving stakeholders in the same way into efficiency analysis as in several new evaluation approaches that embrace the complexity, subjectivity and unpredictability of social interventions. Our assessment is summarised in table 4g and explained in more detail in the subsequent text.

Table 4g. Analytic power and analysis requirements for stakeholder-driven approaches

Stakeholder-Driven Approaches: Analytic Power	
What is the method's analysis level?	Level 1 (identifies efficiency improvement potential within an aid intervention)
To what degree is the method clearly and unambiguously defined?	There is little or no established guidance
How independent is the analysis result from the choice of a specific evaluator under *ceteris paribus* conditions?	Different evaluators can come to entirely different conclusions
How participatory is the method?	Stakeholders participate both in defining analysis criteria and in their valuation

Stakeholder-Driven Approaches: Analysis Requirements					
		Qualitative	Numerical	Financial	Monetarisable
Data	Input-level (minimum requirement)	X			
	Output-level (minimum requirement)	X			
	Outcome-level (minimum requirement)	X			
Time	Overall analysis time needed for evaluator	Several weeks to many weeks			
	Overall time requirements for stakeholders	More than a day per stakeholder involved in the assessment			
Skills	Special skills needed	Strong qualitative and mixed-method analysis and moderating skills			

Description. In addition to refinements in classical evaluation methodology, several authors have also advocated altogether new approaches to evaluation that embrace the complexity, subjectivity and unpredictability of social interventions.[86] While these suggested new approaches differ among each other, many are highly participative and include stakeholders not only as sources of information but also for structuring and guiding the evaluation. In these approaches, the role of the evaluator changes from a detached analytic investigator to an informer, moderator and synthesiser. We summarise these new approaches as stakeholder-driven approaches.

We have struggled somewhat in how to describe the way stakeholder-driven approaches tackle the efficiency concept. Several relevant textbooks remain eerily silent on this subject and experts interviewed sometimes seemed to equate the concept of efficiency with the "old way" of doing evaluation, i.e. the way they were trying to improve. Nevertheless, when encouraged to apply the concept of efficiency (and not specific quantitative methods such as Cost-Benefit-Analysis) within their new, stakeholder-driven, evaluation approaches, some experts contributed ideas.

After some discussion, we developed the following description for a stakeholder-driven efficiency assessment approach:

> *Active management of a holistic dialogue process, including decision-makers, on the question of efficiency, examining efficiency in relation to other evaluation issues from a systems perspective, soliciting and taking into account differing per-*

---

[86] See, for example, Patton (2008): Utilization-Focused Evaluation. Thousand Oaks: Sage Publications and Guba; Lincoln (1989): Fourth Generation Evaluation. Newbury Park, Calif.: Sage Publications.

ceptions of and criteria for efficiency by different stakeholders, supported by qualitative and quantitative investigation as suitable.

While remaining generic, this description already shows fundamental differences to other approaches that apply predetermined evaluation frameworks. In stakeholder-driven approaches to efficiency assessment, stakeholders provide not only most of the data, but also determine what data should be considered important and what standards efficiency should be compared to.

Analytic power. Since stakeholder-driven approaches to the assessment of aid efficiency do not represent established approaches, their analytic power is difficult to assess.

In our opinion, this approach is likely to reveal efficiency improvement potential. Therefore we categorise it among the level 1 approaches. It has the advantage of including the stakeholders' own assessments of efficiency-related aspects that may include important dimensions that are not mapped and virtually immeasurable with other approaches in which predefined categories are valuated by an analyst. In addition, the pacticipatory nature of this approach is likely to favour the identification of ways to improve efficiency. These advantages are paid with the price of assessments potentially biased by vested interests, assessments based on insufficient information or assessments based on insufficient conceptual understanding.

Theoretically, a comparison of several alternative interventions is conceivable based on this approach, which would put it into the level 2 analysis category. We however feel that in most cases, stakeholders will only be able to judge a part of an existing intervention, making it a level 1 approach.

General textbooks on stakeholder-driven evaluation approaches exist and provide some general guidance. We have however not found any literature explaining the application of such approaches to the analysis of efficiency.

While stakeholders drive the approach, the evaluator may influence important aspects of the analysis, for example the initially selected group of stakeholders, the information provided to this group and the selection of the initial set of evaluation questions. As moderator, the evaluator has indirect influence on the result of the overall assessment. Therefore, in our own assessment, different evaluators will usually produce different results. This does however not imply that some results are necessarily wrong. One assessment may for example simply look at another part of an intervention than another one.

Analysis requirements. Data requirements for stakeholder-driven approaches depend on the data the evaluator chooses to provide to stakeholders, data and analysis requested or produced by stakeholders and data the evaluator needs to inform his own synthesis. This implies that different types of data may be needed on all levels of an intervention's results chain which makes a clear-cut description of data needs difficult.

As minimum requirement, we estimate that qualitative data is needed on all levels.

Time requirements are high, both for the evaluator and the participating stakeholders. We estimate the work time needs for the evaluator to range from several to many weeks and those for stakeholders to exceed one day. For large or long-term assignments, these estimates may be exceeded considerably.

In order to implement this approach, in addition to basic analytical skills, the evaluator needs to be well acquainted with qualitative and mixed-methods evaluation approaches and to possess excellent moderating skills.

Suggested literature. No documentation specific to this method has been found.

4.3. Methods for Comparing the Efficiency of Entire Interventions (Level 2 Analysis)

In this section, we present efficiency assessment methods capable of comparing entire aid interventions. As described earlier, level 2 analysis results can assist decision-makers in rational decision-making based on the principle of maximising society's welfare.

Most level 2 methods described in this chapter are highly refined, richly documented in textbooks and intensely researched. This is reflected in our treatment of these methods that exceeds the treatment of most level 1 and level 0 methods in detail.

Overall, we have identified six distinctively separate level 2 methods that are presented in the following five sections:

- Two different Methods for Multi-Attribute Decision-Making (section 4.3.1);

- Cost-Benefit Analysis (section 4.3.2);

- The Effects Method (section 4.3.3);

- Cost-Effectiveness Analysis (section 4.3.4); and

- Cost-Utility Analysis (section 4.3.5).

4.3.1. Methods for Multi-Attribute Decision-Making

Methods for Multi-Attribute Decision-Making (MADM) are powerful and very generally applicable approaches that allow comparing the efficiency of interventions from the perspective of a single decision-maker. MADM methods guide real-world decision-making processes by mapping all options and valuing them according to their perceived utility. In this way, they can be used as an extension of other methods (by incorporating their analysis results into the assessment) or by directly valuing options based on whatever information can be made available. Various MADM methods with differing degrees of scientific rigour exist. In this report, we portray intuitive scoring models and scientific decision analysis as examples for simple and sophisticated approaches. Our assessment is summarised in table 4h and explained in more detail in the subsequent text.

Table 4h. Analytic power and analysis requirements for MADM methods

MADM methods: Analytic Power			
What is the method's analysis level?	Level 2 (assesses the efficiency of an aid intervention in a way that it can be compared with alternatives or benchmarks)		
To what degree is the method clearly and unambiguously defined?	Clear and established analysis procedures exist		
How independent is the analysis result from the choice of a specific evaluator under *ceteris paribus* conditions?	Different evaluators (different decision-makers) can come to entirely different conclusions		
How participatory is the method?	Stakeholder input is restricted to data gathering along established analysis criteria (apart from the decision-maker)		
MADM methods: Analysis Requirements					
---	---	---	---	---	---
		Qualitative	Numerical	Financial	Monetarisable
Data	Input-level (minimum requirement)	X			
	Output-level (minimum requirement)				
	Outcome-level (minimum requirement)	X			
Time	Overall analysis time needed for evaluator	About a day or less to several days (for intuitive scoring models) Many weeks (for scientific decision analysis)			
	Overall time requirements for stakeholders	A few hours or less per stakeholder involved in the assessment			
Skills	Special skills needed	Decision theory and utility theory (for scientific decision analysis)			

Description. Before describing methods for Multi-Attribute Decision-Making, we briefly illustrate the motivation for these approaches: when faced with two or more options and limited resources, decision-makers require a way to assess their preferences and come to a decision.

> *A typical situation is the choice between two alternatives with insufficient information, for example:*
>
> o *Project A has a certain anticipated positive net benefit, for example determined in the course of a Cost-Benefit Analysis (CBA). However, because of analytic challenges, negative environmental effects of project A were not included in the CBA and are described separately in qualitative terms.*
>
> o *Instead, project B has undergone a complete CBA that includes environmental effects and yields a somewhat smaller net benefit than project A.*
>
> *Obviously, the decision-maker is faced with a difficulty. If the negative environmental effects were included into the project A's CBA, the resulting net benefit would be lower. But would it be lower than in project B? The CBAs of projects A and B do not provide sufficient information for that assessment.*
>
> *The decision-maker might also wish to add additional decision-making criteria, such as the political value of each option or the degree of trust he or she places into the capabilities of the evaluators that prepared the analyses.*

In these cases, the decision-maker needs to somehow value different alternatives against each other that may be incompletely described.

Often, such decisions are taken without reference to any explicit methodology. Albert Hirschman describes this situation as follows:

But the generally accepted notion appears to be that decision making on projects involves two, and only two, wholly distinct activities: ascertaining the rate of return and, then, applying feel, instinct, "seat-of-the-pants" judgment, and the like. In actual fact, these latter categories have been left to control a very large portion of the decision-making process.[87]

Multi-Attribute Decision-Making (MADM) methods provide frameworks to guide this type of decision-making, i.e. the assessment, weighing of preferences and comparison of alternatives from the perspective of an individual. Preferences from this person's perspective are systematically recorded in terms of perceived utilities for inputs and results associated with alternative interventions. For simplicity, we refer to this person as the decision-maker.

MADM approaches do not replace analysis of inputs and results, but rather complement them. They can and should build on all available data, including the results of other analysis methods not restricted to efficiency and the opinions and preferences of others. All of this information is needed to describe alternatives as clearly and concisely as possible so that an informed assessment can be made. The advantage of MADM methods lies in the fact that they do not require this information to be in any specific format.

An evaluator can have two roles when MADM approaches are being used by decision-makers. On the one hand, the evaluator can guide a decision-maker through the MADM process by preparing and facilitating the analysis. On the other hand, the evaluator can conduct the assessment from his or her own perspective.

Many different MADM methods exist. One of the oldest applications, a simple procedure for dealing with trade-off decisions, is described as early as 1772 in a letter written by one of the then future founding fathers of the United States, Benjamin Franklin that is reproduced in figure 4a.

---

[87] Hirschman (1995): Development Projects Observed. Washington, D.C.: Brookings Institution, p. 8.

Figure 4a. Historic example for Multi-Attribute Decision-Making[88]

London, September 19, 1772

Dear Sir,

In the Affair of so much Importance to you, wherein you ask my Advice, I cannot for want of sufficient Premises, advise you what to determine, but if you please I will tell you how.

When these difficult Cases occur, they are difficult chiefly because while we have them under Consideration all the Reasons pro and con are not present to the Mind at the same time; but sometimes one Set present themselves, and at other times another, the first being out of Sight. Hence the various Purposes or Inclinations that alternately prevail, and the Uncertainty that perplexes us.

To get over this, my Way is, to divide half a Sheet of Paper by a Line into two Columns, writing over the one Pro, and over the other Con. Then during three or four Days Consideration I put down under the different Heads short Hints of the different Motives that at different Times occur to me for or against the Measure. When I have thus got them all together in one View, I endeavour to estimate their respective Weights; and where I find two, one on each side, that seem equal, I strike them both out: If I find a Reason pro equal to some two Reasons con, I strike out the three. If I judge some two Reasons con equal to some three Reasons pro, I strike out the five; and thus proceeding I find at length where the Ballance lies; and if after a Day or two of farther Consideration nothing new that is of Importance occurs on either side, I come to a Determination accordingly.

And tho' the Weight of Reasons cannot be taken with the Precision of Algebraic Quantities, yet when each is thus considered separately and comparatively, and the whole lies before me, I think I can judge better, and am less likely to take a rash Step; and in fact I have found great Advantage from this kind of Equation, in what may be called Moral or Prudential Algebra.

Wishing sincerely that you may determine for the best, I am ever, my dear Friend,

Yours most affectionately

B. Franklin

Modern MADM approaches range from intuitive weighing exercises to rigorous scientific approaches that are based on multi-attribute utility theory. We present first a simple and then a rigorous version:

- *Intuitive Scoring Models.*[89] All relevant attributes that may contribute to a decision are listed. These can be qualitative, numerical or monetary in nature.

  In the above example of choosing project A or project B, these could be monetary results of the Cost-Benefit Analyses and qualitatively described environmental damage but also other information such as the perceived professional quality of the evaluator or the concordance of the entire project with political priorities. For each attribute,

---

[88] Original source: Franklin; Labaree; Bell; Poor Richard Club (Philadelphia Pa.); Rogers; Benjamin Franklin Collection (Library of Congress); Pforzheimer Bruce Rogers Collection (Library of Congress) (1956): Mr. Franklin : a Selection from his Personal Letters. New Haven, London: Yale University Press; Geoffrey Cumberlege, Oxford University Press, text copied from www.procon.org/view.background-resource.php?resourceID=1474, visited on 27 May 2010 by the author.

[89] These models are usually referred to as „Nutzwertanalyse" in German literature, for example in Zangenmeister (1970): Nutzwertanalyse in der Systemtechnik. München: Verlagskommission Wittmannsche Buchhandlung, Hoffmeister (2008): Investitionsrechnung und Nutzwertanalyse. Berlin: Berliner Wissenschafts-Verlag or in *Arbeitsanleitung Einführung in Wirtschaftlichkeitsuntersuchungen*, H 05 05, VV-BHO – Anhang zu § 7 (Arbeitsanleitung), Vorschriftensammlung Bundesfinanzverwaltung, Berlin, 2001.

weights are developed that reflect their relative importance as perceived by the decision-maker.

Then, the decision-maker assesses the options based on available information and on his or her own preferences by giving a score for each of the attributes, e.g. on a numerical scale from zero (worst) to ten (best). These scores are then weighted based on the weights defined before and added up to total scores for each alternative.

Proponents of this simple approach claim that both the result and the process are helpful. The result guarantees that all attributes considered relevant are taken into account and the process of selecting and weighting attributes can surface considerations that would otherwise not be taken into account.

Critics are worried about the quality of the scoring process on a pseudo-cardinal scale and either suggest more refined methods for converting ordinal feedback into cardinal numbers or point to scientific methods for deriving multi-attribute utility functions as described in the rigorous approach below. Critics also question the simple additive approach that excludes conditional decisions, uncertainty and chance.

In addition, for scoring models, no strong preference is expressed in literature as to whose preferences are actually recorded. It can be the decision-maker, as in our description. However, sometimes, valuations by groups of people that are not further specified are suggested or even recommended. This touches once more on the fundamental issue of interpersonal comparisons of preferences that was described in section 2.2.2. Here, we assume that the decision-maker conducts the entire assessment. If, instead, stakeholder groups contribute to the assessment, this approach becomes similar to the comparative ratings by stakeholders presented in section 4.2.4.

- *Scientific decision analysis.* A more refined and scientifically grounded MADM approach is based on utility theory and can be summarised by the following five steps:[90]

  o *Pre-analysis*, in which the decision-maker and the problem and various action alternatives are identified;

  o *Structural analysis*, in which a decision tree is constructed for the problem at hand;

  o *Uncertainty analysis*, in which probabilities for chance events along the decision tree are analysed;

  o *Utility or value analysis*, in which the decision-maker assigns cardinal utilities along each path in the decision tree;

  o *Optimisation analysis*, in which the path that maximises the expected utility is identified.

This approach builds on a decision tree and therefore allows for multi-layered decisions. Along this tree, the approach also introduces chance nodes to reflect bifurcations not under control of the decision-maker. In this way, chance events are included.

---

[90] Keeney; Raiffa (1993): Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge England; New York, NY, USA: Cambridge University Press, pp. 5.

Finally, utilities for each option along the decision tree are measured by a mental lottery that is often referred to as *standard gamble* and represents one of the founding elements of utility theory. The standard gamble method consists in finding the point where a decision-maker becomes indifferent between a certain outcome X and a gamble in which he or she would have a chance p to "win" the best outcome and a chance 1-p to loose and obtain the worst outcome. The value of p for which a decision-maker becomes indifferent between the certain and the gamble option indicates the relative utility of the certain option.

> *To illustrate this, consider a certain outcome of winning 50 euros and a gamble of winning 100 euros with probability p (and winning nothing with probability 1-p). At p = 50 per cent, the statistical expectation value of the gamble is exactly that of the certain outcome.*
>
> *Now the probability p for which a person becomes indifferent between the certain outcome and the gamble is identified.*
>
> *Risk-averse decision-makers have a probability p higher than 50 per cent, reflecting their preference for certain outcomes, while risk-preferring decision-makers have a probability p smaller than 50 per cent, reflecting their inclination towards risk-taking.*

Until now, we have essentially considered utilities with single attributes. This one-dimensional approach can however be generalised to multi-attribute utilities.

For efficiency assessments, scientific decision analysis is used to obtain utilities in more complex situations, for example to assess the losses in utility for various stages of illness, reflecting the fact that the standard gamble approach is traditionally most applied in the field of health economics. We return to this approach when presenting Cost-Utility Analysis in section 4.3.5.

Proponents of this approach praise its scientific rigour. Critics argue that the standard gamble model is difficult to apply meaningfully in practice.

Analytic power. MADM methods are designed to facilitate choosing between alternatives. They can be used to assess the efficiency of entire aid interventions and are therefore classified as level 2 analyses.

It should be noted, however, that this assessment is carried out from the perspective of a single person and that efficiency is understood in terms of maximising total utility from the perspective of this person for different consequences, i.e. efficiency is understood as maximisation rather than transformation efficiency (see section 2.2).

The strength of MADM approaches lies in the fact that they introduce a systematic and transparent framework into the decision-making process. While all other approaches presented in this report provide information for consideration by decision-makers, MADM approaches walk decision-makers through their assessment.

Between the two approaches presented, obviously, the scientific decision theory approach is more sophisticated. It can handle chance and uncertainty and map conditional utilities in several layers. The assessment of utility itself builds on a solid theoretical foundation. Therefore, we consider this approach more reliable.

Both approaches require descriptive information, data and analysis so that the decision-maker can express his or her utilities for each scenario. MADM approaches do not replace other analyses, including other analyses portrayed in this report. Instead, they build on it.

Literature on both approaches presented in this section exists, even going into considerable theoretical detail.

Regarding the degree of subjectivity, the evaluator has some leeway in selecting valuation categories and in choosing and designing the data collection mechanism. More importantly, however, is the process in which utilities are determined. This is done from the perspective of a single person, either a decision-maker (if the evaluator acts as facilitator) or the evaluator. Even if this assessment is informed by convincing data and analysis results, the utility valuation conducted by the decision-maker reflects his or her preference structure and therefore remains subjective. This implies that different decision-makers – all other things equal – could come to entirely different decisions.

Stakeholder feedback is a likely source of information for the decision-maker. It can be obtained directly or indirectly, i.e. as part of further analysis. We estimate that in most cases, stakeholder feedback is restricted to information gathering along predefined categories.

Analysis requirements. MADM methods require that different options are described in some way but do not specify how. Ralph Keeney and Howard Raiffa describe this as follows:

> It is immaterial at this point what the underlying scales of these x's [the conse-quences] are. Each x could be a scalar, a vector, or a paragraph of prose de-scribing this consequence.[91]

Therefore, we consider qualitative data on the input and outcome-level as minimum require-ments. However, the quality of the assessment increases with the richness with which op-tions can be described.

In terms of time and skill requirements, the two MADM methods presented in this section differ considerably. We therefore provide different estimates for each method.

Intuitive scoring models can be applied in a "quick and dirty" manner requiring less than a day's combined time for the decision-maker and an evaluator who guides the process. Even if applied with more rigour, the approach should not require more than a few days.

Instead, utility theory-based approaches require considerable time. Conducting, sometimes iteratively, all analysis steps of scientific decision analysis may, in our opinion, require many weeks of combined work time for the decision-maker and an evaluator.

For scoring approaches, only basic analytical and interviewing skills are needed. It is crucial, however, that the evaluator has a clear understanding of the implications of approximations made, for example when generating cardinal data from ordinal values.

Utility theory-based approaches, instead, require a sound knowledge of decision analysis and utility theory.

---

[91] Ibid., p. 132.

Good practice and suggestions for extension of the basic design

- MADM approaches should be viewed as complement and not as alternative to other analysis approaches. The strength of MADM approaches is their ability to provide a structure for the decision-making process that is highly flexible in terms of data requirements.

- If conducted by an evaluator instead of a decision-maker, the evaluator needs to present his or her reasoning when assigning utilities to different options in order to allow those responsible for making decisions to confirm or adapt the conclusions drawn by the evaluator.

Suggested literature

- Yoon; Hwang (1995): Multiple Attribute Decision Making : an Introduction. Sage University Papers Series. Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage Publications 07.

- Keeney; Raiffa (1993): Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge England; New York, NY, USA: Cambridge University Press, and further references therein.

- Zangenmeister (1970): Nutzwertanalyse in der Systemtechnik. München: Verlagskommission Wittmannsche Buchhandlung.

- Hoffmeister (2008): Investitionsrechnung und Nutzwertanalyse. Berlin: Berliner Wissenschafts-Verlag.

- Arbeitsanleitung Einführung in Wirtschaftlichkeitsuntersuchungen (2001): H 05 05, VV-BHO – Anhang zu § 7 (Arbeitsanleitung), Berlin.

4.3.2. Cost-Benefit Analysis

Cost-Benefit Analysis (CBA) expresses all costs and benefits associated with an aid intervention in monetary terms. In principle, this analysis can cover all social and private, direct and indirect and tangible and intangible inputs and results. Since both costs and benefits are expressed in the same unit, net benefits can be explicitly calculated, making CBA a useful basis for decision-making.

CBA is the most widely applied and a very thoroughly researched level 2 analysis. It has many fervent proponents but also some ardent critics. Our assessment is summarised in table 4i and explained in more detail in the subsequent text.

Table 4i. Analytic power and analysis requirements for Cost-Benefit Analysis

Cost-Benefit Analysis: Analytic Power	
What is the method's analysis level?	Level 2 (assesses the efficiency of an aid intervention in a way that it can be compared with alternatives or benchmarks)
To what degree is the method clearly and unambiguously defined?	Clear and established analysis procedures exist
How independent is the analysis result from the choice of a specific evaluator under *ceteris paribus* conditions?	Results obtained by different evaluators are expected to be very similar (if key assumptions are predefined) or may vary somewhat (if key assumptions are made by evaluator)
How participatory is the method?	Stakeholder input is restricted to data gathering along established analysis criteria

Cost-Benefit Analysis: Analysis Requirements					
		Qualitative	Numerical	Financial	Monetarisable
Data	Input-level (minimum requirement)				X
	Output-level (minimum requirement)				
	Outcome-level (minimum requirement)				X
Time	Overall analysis time needed for evaluator	Several weeks to many weeks			
	Overall time requirements for stakeholders	A few hours or less per stakeholder involved in the assessment			
Skills	Special skills needed	Advanced economic analysis			

Description. Cost-Benefit Analysis (CBA) valuates costs and benefits from a society-wide perspective or from the perspective of groups within that society in monetary terms.[92]

The procedure for conducting CBA shows similarities with microeconomic investment analysis described in section 4.2.3. The principal difference lies in the relative scope of both methods: CBA describes social and private, direct and indirect and tangible and intangible inputs and results and therefore aims at recording all welfare effects associated with a development intervention. Financial analysis, instead, is limited to financial quantities.

In CBA, all inputs and results associated with a development intervention are expressed by monetary units for each year the intervention covers. Costs and benefits of different years are then discounted to their value in one point in time, a procedure identical to calculating present values (by discounting cash flows occurring in different years) in financial analysis. Finally, the net present benefit (usually simply referred to as net benefit) at one point in time can be calculated by subtracting total present costs from total present benefits.

In this way, the total net benefits of an intervention can be calculated and expressed as a single number. CBA is the only analysis method presented in this report that *seems* to make absolute welfare statements, i.e. that allows us to determine whether an intervention is beneficial or disadvantageous to society without comparing it to alternatives. In reality, however, also CBA makes only relative statements. When calculating opportunity costs and when determining the value of the discount rate to be used, we implicitly compare to alternative scenarios.

---

[92] Strictly speaking, any other (non-monetary) unit would do as well if all costs and benefits could be expressed in that unit.

Instead of reporting net benefits, results of CBA are often expressed as benefit-cost ratios[93] or as economic rates of return. This has certain advantages and disadvantages as explained below.

CBA is applicable to interventions that change the allocation of resources as, for example, infrastructure investment projects (e.g. roads and dams) and social service programmes (e.g. education and health).

Since CBA incorporates the idea of trade-offs between different members of society that was first introduced as part of the Kaldor-Hicks concept,[94] it is, *a priori*, less suitable for interventions that mainly aim at redistributing costs and benefits, such as tax reform and transfer programmes. If the analysis is conducted from the perspective of the entire society, distributional effects may remain invisible if the winners' positive benefits and the losers' negative benefits cancel each other out.[95] This limitation can however easily be remedied by conducting additional CBAs from the perspective of the winning and losing groups.

CBA has its fervent proponents and ardent critics, more so than any other method presented in this report. Many economists consider it the gold standard of efficiency analysis while others feel that it does not produce useful results outside of a very limited application radius.

For example, Franck Wiebe, chief economist at the Millennium Challenge Corporation, comments:

> *Aid interventions need to positively prove their development efficiency. Given the poor track record of aid effectiveness, I feel that it is better to shift resources towards investments and activities that have solid evidentiary basis of expected impact and away from those that do not (or to limit the funding of those without evidence to smaller, exploratory or pilot activities designed to generate such evidence).*

Michael Patton, author of many evaluation textbooks and former president of the American Evaluation Society sees things quite differently:

> *In my judgment, cost-benefit analysis fits within a very narrow and constrained development space where major variables are known, agreed on, predictable, and controllable. That space is rare. To expand that space, economists make assumptions, typically over-simplifying, thus treating complexity as simple. This over-simplification distorts the complex realities and dynamics of development. It becomes largely an academic exercise. The assumptions econometricians are willing to make so that their equations will work boggle the mind.*

Overall, CBA is one of the most frequently applied and at the same time most sophisticated efficiency assessment methods. Its sophistication and the transparency on assumptions and approximations caused by its structured approach represent important strengths.

At the same time, this sophistication and transparency allows for an explicit discussion of a number of particularities and potential pitfalls when conducting CBA that is presented in some detail below. Such a discussion is not possible for most other methods presented in

---

[93] Benefit-cost ratios can be calculated dynamically, i.e. based on separately discounted costs and benefits or statically, based on benefits and costs for one specific year.

[94] See section 2.2.1.

[95] In real markets, allocation and distribution effects may be coupled, leading to net welfare effects of what we refer to as distributional effects.

this report, simply because those methods lack the degree of sophistication encountered in CBA.

Before presenting these points, we encourage the reader to keep in mind that the absence of similar discussions for other methods does not imply that those are free of similar – or more serious – particularities and weaknesses. We restrict our discussion to a number of aspects that either seem of great importance or seem not homogeneously reflected in standard litera-ture. We also do not attempt to provide a step-by-step instruction for CBA or to comprehen-sively summarise the rich literature on methodological subtleties that fill entire textbooks. Both can easily be found elsewhere.[96]

> *Determination and monetarisation of costs and benefits*. In Cost-Benefit Analysis, benefits can be approximated by willingness to pay, i.e. the maximum amount of money a person would be willing to pay for a desired result or to avoid an undesired result. Total benefits are then computed by summing up individual benefits. Costs re-lated to an intervention are computed as *opportunity costs*, i.e. the value of the bene-fits forgone when using inputs for the intervention at hand and not for other purposes. However, in many cases, budget costs represent a valid approximation.[97] This proc-ess usually assumes, implicitly or explicitly, some type of social welfare function as described in section 2.2.3.
>
> Under special circumstances, i.e. for isolated perfect markets, benefits can be calcu-lated straightforwardly based on the willingness to pay. In such a market, all consum-ers that are willing to pay less than the actual market price simply don't buy the prod-uct or service but those consumers that are willing to pay more than the market price enjoy a benefit that is called *consumer surplus*: they obtained something for less money than they are prepared to pay. Similarly, producers that would have been will-ing to sell for less also enjoy a benefit, the *producer surplus*.
>
> In other cases, market imperfections make market prices a poor basis for the deter-mination of benefits. Typical examples for market imperfections are externalities and natural monopolies.
>
> In such cases, artificially constructed prices, so called *shadow prices*, are utilised that aim at describing true willingness to pay. For example, if the price of a good does not include the damage to the environment its production causes, an analyst might use a shadow price that is higher than the market price in order to internalise this negative benefit into the analysis. Today, the probably most accepted shadow pricing method-ology for aid projects was first introduced by Ian Little and James Mirrlees in 1974.[98]

---

[96] An excellent step-by-step summary can be found in the last chapter of Gramlich (1990): A Guide to Benefit-Cost Analysis. Englewood Cliffs, N.J.: Prentice-Hall, pp. 223. For more in-depth reading we recommend Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall. More literature is provided at the end of this section.

[97] See, for example, Gramlich (1990): A Guide to Benefit-Cost Analysis. Englewood Cliffs, N.J.: Prentice-Hall, pp. 227 and Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, pp. 93.

[98] Basic shadow price methodology for aid projects was developed in the 1970s by UNIDO (see UNIDO (1972): Guidelines for Project Evaluation. New York: UNIDO) and by I.M.D. Little and J.A. Mirrlees (see both the original publication: Little; Mirrlees (1974): Project Appraisal and Planning for Developing Countries. New York: Basic Books, and the sequel: Little; Mirrlees (1994): The Costs and Benefits of Analysis: Project Appraisal and Planning Twenty Years on. Cost-Benefit Analysis. Layard and Glaister. Cambridge: Cambridge University Press). Both approaches were synthesised by L. Squire and H.G. van der Tak (see Squire; Tak (1975): Economic Analysis of Projects. A World Bank

Experts today seem to largely agree on the principles of application and determination of shadow prices but differ on the level of practical details.[99]

In some cases, complications occur if secondary markets are affected by the results of an intervention. Typical examples are substitution effects, such as the replacement of food crops by bio fuel farming or distortion effects on prices in secondary markets. If secondary effects exist, these are either considered explicitly by application of general equilibrium models or approximated.[100] The European Commission's Guide to Cost-Benefit Analysis of Investment Projects[101] summarises this as follows:

> *Because, however, the direct computation of the shadow prices by a general equilibrium model of the economy is constrained by lack of data, computation shortcuts have been proposed. The most famous one is the Little-Mirrlees 'border price rule' for traded goods and the 'long run marginal cost rule' for non-tradable goods. In fact, applied cost-benefit analysis needs always to deal with data limitations and the choice of a partial versus a general equilibrium approach is in fact a matter of convenience.*

In special cases, entirely different approaches are needed to estimate benefits that are not traded at all.[102] A typical example is the valuation of a human life. Several methods exist, all of which are rough approximations in a mindboggling valuation exercise. For example, the *Discounted Future Earnings (DFE) method*, often used in lawsuits, calculates the present value of a person's future earnings and the *Required Compensation (RC) approach* extrapolates the benefit value of a human life from increasing salary risk premiums for increasingly hazardous jobs.

Another special example concerns the valuation of public goods, for example a public wilderness area. A lower boundary for visitors' benefits can be established by the *travel cost method*, i.e. the amount of money spent for travelling to the area. If the area were destroyed as a consequence of an intervention, this method would yield a minimum boundary valuation of the perceived environmental damage caused.

For these special examples, the valuation of benefits may become very tricky and analysts need to be careful not to lose sight of the realism of their estimates as ironically expressed by Carol Weiss:

Research Publication. Baltimore: Published for the World Bank by Johns Hopkins University Press) who emphasised the Little/Mirrlees approach. With some modifications, this lead to today's widely accepted LMST accounting method for calculating shadow prices for aid projects.

[99] For example, Boardman et al describe this as „*Surprisingly, perhaps, there is not only agreement among the experts on using shadow prices, but also on the basic methods to use in determining the values of these shadow prices. Nonetheless, considerable variation remains in the details on how these methods are applied in practice.*" (cited from: Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, p. 441).

[100] For an introduction to general equilibrium models see, for example, Gramlich (1990): A Guide to Benefit-Cost Analysis. Englewood Cliffs, N.J.: Prentice-Hall, pp. 79 or Chowdhury; Kirkpatrick (1994): Development Policy and Planning : an Introduction to Models and Techniques. London ; New York: Routledge.

[101] European Commission (2008): Guide to Cost-Benefit Analysis of Investment Projects. Brussels: European Commission, p. 46.

[102] For a description of some of these methods see, for example, Gramlich (1990): A Guide to Benefit-Cost Analysis. Englewood Cliffs, N.J.: Prentice-Hall, pp. 66 and pp. 134.

> *To move from evidence of outcomes to dollar benefits, analysts often posit theories of amazing grandeur.[103]*

Overall, for most cases in which direct, tangible and expected benefits describe the effects of an intervention well, the valuation of costs and benefits can be considered a well-established standard procedure. Under special circumstances, there is however an invisible line beyond which approximations lose their practical value. What should be done under those circumstances, beyond that line? Edward Gramlich suggests the following:

> *We have talked about [valuing gains when there are] direct markets and allied markets, but not so much about what to do when there are no markets. One would be hard-pressed to come up with any market to use to value the MX missile, the value to reduce roadside litter for trash pickup or bottle return bills, the value of reduced emission of sulfur dioxides or chlorofluorocarbons, and on down the list. What to do then?*

> *It sounds paradoxical, but the best approach is to do nothing at all. Benefit-cost analysis is often criticized for undervaluing things that cannot be valued, but if done properly, that criticism is wide of the mark. If things cannot be valued by any known technique, just be honest about it and keep that limitation up front. […]*

This leads to the simple insight that Cost-Benefit Analysis is applicable as long as the necessary approximations allow for a reasonably realistic representation of true costs and benefits.

An additional remark seems appropriate here. In some instances, interventions may have vaguely defined and unclear objectives. In these situations, the determination and subsequent valuation of costs and benefits is difficult since, *a priori*, there is little guidance in what these could be.

- *Taking time and chance into account.* Future costs and benefits need to be discounted in order to obtain their value in today's dollars, i.e. their present value. This is a straightforward generalisation of the observation that for most people receiving 100 US Dollars a year from now is less attractive than receiving 100 US Dollars today.[104]

The choice of the discount rate has a large impact on the final result, both in terms of the absolute and the relative values of net benefits.

  - In absolute terms, consider an intervention with a constant net benefit stream of 1,000 euros per year that continues forever. In this case, the present net benefit formula is simply given as the annual net benefit divided by the discount rate. This implies a net present benefit of 20,000 euros using a five per cent discount rate but net benefits of 10,000 and 100,000 euros for discount rates of ten per cent or one per cent, respectively, illustrating the sensitivity of net present benefits on the discount rate chosen.

---

[103] Weiss (1998): Evaluation: Methods for Studying Programs and Policies. Upper Saddle River, N.J.: Prentice Hall, pp. 248.

[104] Putting 100 US Dollars into a bank account with, for example, an annual interest rate of three per cent (and no fees) would yield 103 US Dollars in a year's time. Reciprocally, the value of receiving 100 US Dollars a year from now could then be estimated to be equivalent to receiving 97.1 US Dollars today.

- o In relative terms, high discount rates favour interventions that generate quick benefits and incur costs later, simply because, in this case, benefits are discounted less than costs.

Regarding the choice of specific discount rates, experts disagree somewhat on both the method and the rate to use when discounting future costs and benefits. As Anthony Boardman et al. conclude after describing different approaches:

> *There has been considerable debate as to the appropriate method of discounting, as well as the specific estimate for the SDR [Social Discount Rate].*

> *[...] Our discussion of the issues should have made it clear that analysts are unlikely ever to have complete confidence in whatever discounting procedure they use. Even in the case of the more simple discounting methods, choice of the appropriate social discount rate requires judgement.[105]*

Most experts agree that it is technically wrong to use different discount rates in a single decision-making context. Strictly speaking, this is only valid as long as risk and the time intervals considered in all alternatives are the same.

In practice, different discount rates are used by different agencies and sometimes within one agency for different interventions. Some experts also feel that applying standard discounting techniques for effects with long-term consequences may lead to erroneous results.[106]

Obviously, discount rates need to be selected and justified in a transparent manner by evaluators and the choice of discount rate needs to be considered when comparing CBA results.

If different periods are considered, different discount rates may be appropriate for each period.

In addition, uncertainties about future events may be included by calculating the expected values for each possible outcome, i.e. weighing each outcome with its respective probability. If opportunities or risks are very large, additional adjustments may be required.[107]

- *Net present benefits versus benefit-cost ratios and economic rates of return.* In most appraisal and evaluation reports reviewed that contained CBAs, results were expressed by the Economic Rate of Return (ERR) and sometimes by benefit-cost ratios. Often, values for net benefits were not explicitly provided. In the following paragraphs we argue that, instead, the net present benefit criterion should be used when applying CBA.

---

[105] Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, p. 269.

[106] See, for example, Price (1996): Discounting and Project Appraisal: From the Bizarre to the Ridiculous. Cost-Benefit Analysis and Project Appraisal in Developing Countries. Kirkpatrick and Weiss. Cheltenham, UK; Brookfielt, Vt., US: E.Elgar pp. 90.

[107] See, for example, Gramlich (1990): A Guide to Benefit-Cost Analysis. Englewood Cliffs, N.J.: Prentice-Hall pp. 73, or Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, pp. 165.

The Economic Rate of Return is defined as the discount rate for which net present benefits are zero. ERRs do not require analysts to make decisions about what discount rates to use, which is often considered an advantage. Benefit-cost ratios simply divide all (present) benefits by all (present) costs. In some cases, the ratio is inverted (cost-benefit ratio) or special ratios covering only some benefits or some costs are considered.[108]

Compared to net benefits, these quantities provide an impression of how much benefit is produced per unit of cost. Many practitioners feel that these quantities are better suited for the selection of interventions since they normalise benefits to one unit of cost and render interventions of different sizes comparable. Some aid agencies that regularly conduct CBAs have established cut-off values for Economic Rates of Return below which interventions have to provide additional justification for their merit.

The frequent use of Economic Rates of Return and benefit-cost ratios stands in contrast to the academic literature that advises against their use. Instead, net present benefit criteria (see section 2.6) are advocated as selection principles. There are several reasons for this:

- First, selecting projects on the basis of their benefit-cost ratios or ERRs may not lead to a net benefit maximising portfolio of interventions because of three reasons:

    - Even based on the same initial investments, interventions with fast benefit-payback may produce larger ERRs, but lower total net benefits than interventions for which benefits occur later after the initial investment which may lead to inverse ranking along net benefits versus ERRs.[109]

    - Interventions may be "lumpy", i.e. not divisible or scalable. If we worked down a list where interventions are ordered according to their ERR, we might end up with unused resources. In this context, it is possible that another combination of interventions would exploit available resources more completely and produce superior net benefit.

        A simple example is that of a large intervention that would be the next candidate in the ERR-ordered list but that cannot be implemented because its resource requirements slightly exceed the available, remaining resources. According to the ERR rule, no further interventions would be implemented even if other interventions (further down the list)

---

[108] For a detailed discussion of the applicability of cost-benefit ratios for decision making in a series of scenarios see: Fuguitt; Wilcox (1999): Cost-Benefit Analysis for Public Sector Decision Makers. Westport, Conn.: Quorum, pp. 81.

[109] *Imagine 2 projects with the same initial investment of 100,000 US Dollars. Project A (the "fast payback" project), produces 150,000 US Dollars worth of net benefits after one year and nothing after that, while project B (the "slow payback" project), produces a one-time net benefit of 300,000 US Dollars five years after the initial investment. This leads to the following Net Present Values (NPVs) and Economic Rates of Return (ERRs), if a discount rate of 10 per cent is assumed:*

- *NPV of 36 thousand US Dollars and an ERR of 50.0 per cent for project A (fast payback); and*

- *NPV of 86 thousand US Dollars and an ERR of 24.6 per cent for project B (slow payback).*

Following the net benefit rule, intervention A should be selected. Instead, selecting the intervention with the higher ERR leads to net benefits that are about 50,000 US Dollars lower.

would have lower resource requirements and therefore be implement-able.

  ▪ Benefit-cost ratios and ERRs provide a sense of how much benefit is produced per unit of cost. The cost calculated in CBA, however, almost never matches the constraints that force agencies to select among projects with positive net benefits. Such constraints are, for example, out of the pocket expenditures or staff capacity of an aid agency or implementing partners. Instead, costs in a Cost-Benefit Analysis may or may not reflect these constraints but contain also other less constrained or unconstrained quantities.

Edward Gramlich summarises his view on benefit cost ratios and ERRs as follows:

> *As with the benefit-cost ratios [...] the [...] rate of return often gives correct information about a project, but sometimes not. It is either redundant or wrong. The most sensible policy seems then not even to compute it, but rather to focus on the net benefits of a project, which is always the proper indicator.* [110]

However, without budgetary or other constraints all measures can be used interchangeably, as long as the same reference discount rate is used.

In this case, if we followed the net benefit rules described in section 2.6, all interventions with net benefits larger than zero should be implemented. These interventions can interchangeably be characterised by their benefit-cost ratios being larger than one or by their economic rate of return being larger than the usual discount rate.

o Second, the net present benefit is immune to shifting negative benefits into costs and vice versa. In contrast, benefit-cost ratios depend on how inputs and results are distributed into benefits and costs. Boardman et al describe this as follows:

> *Furthermore, the benefit-cost ratio is sensitive to whether negative willingness-to-pay (willingness-to-accept) amounts are subtracted from benefits or added to costs. [...] Thus, benefit-cost ratios are subject to manipulation. For these reasons [this argument and the argument presented above], we recommend that analysts avoid using benefit-cost ratios and rely instead on net benefits to rank policies.* [111]

In the case of net present benefits, only the annual differences between benefits and costs enter the analysis which leads to the same results no matter how negative benefits are assigned.

o Third, multiple ERRs may exist for one intervention, rendering this criterion ambiguous.

---

[110] Gramlich (1990): A Guide to Benefit-Cost Analysis. Englewood Cliffs, N.J.: Prentice-Hall pp. 100. Gramlich uses the term internal rate of return for what we refer to as Economic Rate of Return.

[111] Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, pp. 32. See also pp. 15 of the same textbook.

The mathematical equation determining the ERR is a potential equation of power n+1 (where n is the number of years included into the analysis). This equation has n+1 mathematical solutions and possibly more than one meaningful solution for ERRs or expressed less mathematically: when tuning the discount rate from zero upwards, the net present benefit may pass through zero more than just once.[112]

Interventions characterised by substantial up-front *and* close-down investments may, for example, have dual ERR solutions. Outside of the development field, mining operations and oil platforms are examples.

Concluding our discussion of what quantity is most suitable for the selection of interventions, we observe that Economic Rates of Return and benefit-cost ratios, while more commonly used, are quantities that need to be interpreted with care. If net present benefit considerations are possible, these should be used instead.

- *Necessity of additional analysis of distributional effects.* The maximisation of net benefits alone – from the perspective of the entire society – may not satisfy the goals of some aid interventions. For example, consider two different interventions with the same net present benefits: Intervention A makes poor people somewhat poorer and rich people so much richer that the sum of net benefits (negative for the poor, positive for the rich) are equal to those of intervention B that increases everybody's income somewhat. Based on this analysis of net benefits, both interventions are equivalent but aid donors may have a strong preference for B over A.

  As a reaction, authors of recent CBA textbooks have begun to include additional analysis of distributional effects into the standard description of CBA. Distributional analysis can be done in two different ways.

  - On the one hand, net benefits can be calculated from the perspective of several beneficiary sub-groups in addition to the perspective of the entire society. This yields explicit information about how net benefits are distributed over these groups.

  - On the other hand, a focus on specific target groups can be directly incorporated in CBA by artificially inflating net benefits to those groups with respect to other groups, for example by introducing distributional weights.[113]

  Some aid agencies consulted for this report have already considered it standard CBA practice to conduct the analysis both from a society-wide perspective as well as from the perspective of all relevant groups within that society and thereby effectively providing all required distributional analysis.

  It should be added that non-ideal markets may anyhow prohibit a separate treatment of allocation and distribution because of linkages between both effects.

- *Sensitivity testing.* Most CBA textbooks consider sensitivity analysis to be part of the CBA analysis process. Most reports containing CBA reviewed for this study did however not contain a sensitivity analysis.

---

[112] The *Descartes' rule of signs* helps in determining an upper limit for how many real solutions exist by counting the number of times the sign of the coefficients of the net benefit polynomial changes.

[113] See, for example, Gramlich (1990): A Guide to Benefit-Cost Analysis. Englewood Cliffs, N.J.: Prentice-Hall pp. 115, or Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, pp. 488.

Sensitivity analysis tests the dependence of analysis results on variations in critical assumptions, for example of the discount rate. Often, this additional analysis can be easily conducted based on existing spreadsheets that were built for the determination of net present benefits.

The rationale for conducting sensitivity analysis is to provide an estimate of the error margin of CBA results. Obviously, CBA results are only meaningful if they remain reasonably stable under reasonable variations of assumptions.

- *Presentation of results*. In contrast to many other methods presented in this report, CBA aggregates analysis results into a single efficiency measure. Even if accompanied by additional distributional and sensitivity analyses, CBA results are probably the most concise way for summarising efficiency-related information of an entire aid intervention. This is obviously a great achievement but poses risk at the same time.

In order to correctly interpret CBA analysis results, additional information about underlying assumptions and limitations of the analysis is usually required. Carol Weiss summarises this as follows for economic efficiency analysis in general:

> *As noted, the technique is awash in judgments. With increasing experience many procedures are becoming standardized but analysts still face issues of the perspective they should take, which costs and which benefits to consider, how costs and benefits should be measured, how they should be valued, what groups should be compared, and what discount rate should be applied. Depending on the answers to those questions, results can differ by orders of magnitude.[...] Readers of cost-benefit and cost-efficiency analysis have to read the fine print.*[114]

In reports reviewed for this study, this fine print was often missing, leaving the reader somewhat in the dark regarding important assumptions. In some cases, the fact that environmental effects were neglected altogether was not mentioned or only alluded to in footnotes. Yet, a clear description of basic assumptions and limitations of the analysis is as important as the analysis results themselves. One way out, apart from explicitly describing assumptions and limitations in every report, is to standardise CBA across an agency and use and refer to this standard in reports.

Analytic power. Cost-Benefit Analysis is a powerful level 2 analysis. For interventions that allow for quantification and monetarisation of all costs and benefits, CBA can condense information about the net welfare effects into a single number. In this sense, it is the most comprehensive of all efficiency assessment methods presented in this report.

For interventions with some outcomes that are hard to quantify, CBA can be applied as level 1 analysis, i.e. by comparing the efficiency of those components or aspects of interventions that can be expressed in monetary terms.

In our review of literature, CBA was the best documented efficiency analysis method. Textbooks and academic publications covered even subtle and exotic analysis details. Therefore, or assessment on available guidance is that clear and established analysis procedures exist.

---

[114] Weiss (1998): Evaluation: Methods for Studying Programs and Policies. Upper Saddle River, N.J.: Prentice Hall, pp. 250.

Not all textbooks agree, however, on all details. For example, some textbooks understand that the analysis of distributional effects, the inclusion of chance, and sensitivity testing are integral parts of CBA, while others do not.

Moreover, when applied in practice, assumptions about time horizons, discount rates and the degree of inclusion of indirect or intangible costs and benefits can vary from cases to case and from agency to agency. Some appraisal or evaluation reports we reviewed made no mention of these important assumptions. In these cases, the evaluator has considerable freedom in making assumptions that have influence on the analysis result. The stability of analysis results with respect to different evaluators conducting the analysis therefore depends on how well the analysis procedure is defined, for example by the commissioning:

- If all vital assumptions are specified, results obtained by different evaluators can be expected to very similar.

- If instead, the evaluator is expected to make his or her own assumptions regarding the time horizon, the discount rate, and what categories of costs and benefits to include into the analysis, results obtained by different evaluators may vary.

Stakeholders are usually part of the analysis in the sense that information on costs and benefits is gathered along established analysis dimensions.

Analysis requirements. Cost-Benefit Analysis requires monetary data on inputs and on outcomes. While costs are obviously related to inputs, our choice of characterising benefits by outcome-level data requires some explanation.

Most often, benefits of an intervention are calculated based on the outcomes it produces. Outcomes, for example expressed by numerical data, is then valuated and expressed in monetary terms. This valuation, in principle, implies all further consequences these outcomes may produce but often does not address them explicitly. For example, for interventions aimed at reducing poverty, the analysis may include beneficiaries' household income effects but not explicitly consider welfare effects triggered by these. In order to reflect this, we choose to characterise data requirements on the results side only on the level of outcomes and not of outputs or impacts.

Work time requirements for conducting CBA vary with the complexity of the intervention and with the degree of difficulty in data gathering. For standard CBA applications in which existing operational data can be used and for which established valuation standards exist, the entire analysis can be conducted in a matter of several weeks or less. If data needs to be gathered directly from participants and beneficiaries or if new valuations of hard-to-quantify effects need to be produced, analysis time may be considerably longer. For example, Anthony Boardman et al cite costs for large CBAs conducted by the American Environmental Protection Agency in the 1980s and 1990s in the order of magnitude of one million US Dollars.[115]

We assume that work time requirements for stakeholders will usually not exceed a few hours per stakeholder involved.

Evaluators conducting Cost-Benefit Analysis need to have special training and advanced economic analysis skills.

---

[115] Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, p. 6.

Good practice and suggestions for extension of the basic design. Good practices and extensions of the basic design for CBA were already mentioned in the description of the method. We highlight the following points here:

- *Agreement on and standardisation of key assumptions.* A number of critical key assumptions underlie every CBA, for example the choice of discount rate, the number of years considered in the present value analysis, the assumptions made when valuing costs and benefits and the specific analysis perspective taken.

  On the one hand, these assumptions naturally differ for different types of interventions. On the other hand, analysis results based on different sets of assumptions are hard to compare. Therefore, it is useful to develop standards for these assumptions for different types of interventions.

  One example are guidelines and standards for conducting mandatory Cost-Benefit Analysis for federal public programmes in the US that provide concrete guidance and, for example, establish the discount rate to be used.[116]

- *Clear presentation of key assumptions and limitations of the analysis.* Key assumptions that have noticeable influence on the overall analysis result need to be clearly communicated. A sensitivity analysis is a useful tool to illustrate the effects of reasonable variations in these assumptions. If such variations lead to contradictory results, the analysis result itself has little practical value. Similarly, if important costs or benefits have been excluded from the assessment, for example because they are difficult to quantify, these omissions have to be clearly stated. Ideally, these contributions are then assessed by some other, potentially qualitative, method.

- *Analysis of distributional effects.* CBA from a society-wide perspective allows for trade-offs between winners and losers, which renders transfers of welfare between different groups invisible unless a description of these effects is required. In addition to total net benefit analysis, the additional analysis of net benefits for specific target groups is useful. The analysis of such distributional effects may provide additional important information for decision-making.

  As a good practice example, the Millennium Challenge Corporation (MCC) has recently developed *poverty scorecards* that show incremental present benefit effects for several different groups (defined by their poverty level) and allow comparing relative and absolute effects per beneficiary for each of these groups.

Suggested literature. Although we provide more literature on CBA than on other methods, considerably more relevant literature exists. We merely list the books and publications that we have reviewed for this report.

Textbooks on CBA (newest publications first):

- Mishan; Quah (2007): Cost-Benefit Analysis. London; New York: Routledge.

- Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall.

---

[116] *Memorandum for Heads of Executive Departments and Establishments. Subject: Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs*, Circular No. A-94 Revised (Transmittal Memo No. 64), Office of Management and Budget, Executive Office of the President of the United States, 1992, downloaded on 7 July 2010 from: http://www.whitehouse.gov/omb/rewrite/circulars/a094/a094.html#8.

- Adler; Posner (2006): New Foundations of Cost-Benefit Analysis. Cambridge: Harvard University Press.

- Brent (2006): Applied Cost-Benefit Analysis. Cheltenham, UK ; Northampton, MA: Edward Elgar.

- Campbell; Brown (2003): Benefit-Cost Analysis: Financial and Economic Appraisal Using Spreadsheets. Cambridge, UK ; New York: Cambridge University Press.

- Fuguitt; Wilcox (1999): Cost-Benefit Analysis for Public Sector Decision Makers. Westport, Conn.: Quorum.

- Belli; Anderson; Barnum; Dixon; Tan (1997): Handbook on Economic Analysis of Investment Operations. Washington, D.C.: The World Bank.

- Nas (1996): Cost-Benefit Analysis: Theory and Application. Thousand Oaks, Calif.: Sage Publications.

- Dinwiddy; Teal (1996): Principles of Cost-Benefit Analysis for Developing Countries. Cambridge; New York: Cambridge University Press.

- Gramlich (1990): A Guide to Benefit-Cost Analysis. Englewood Cliffs, N.J.: Prentice-Hall.

Introductions to CBA:

- European Commission (2008): Guide to Cost-Benefit Analysis of Investment Projects. Brussels: European Commission.

- Langhammer; Nägele (2005): Einführung in die Gesamtwirtschaftliche Investitionsanalyse. Frankfurt am Main: KfW Entwicklungsbank.

- Hemmer (2005): Einführung in die Cost-Benefit Analyse. Frankfurt am Main: KfW Entwicklungsbank.

- Kee (2004): Cost-Effectiveness and Cost-Benefit Analysis. Handbook of Practical Program Evaluation. J. S. Wholey, H. P. Hatry and K. E. Newcomer. San Francisco: Jossey-Bass pp. 506.

- Rossi; Lipsey; Freeman (2004): Evaluation : a Systematic Approach. Thousand Oaks, CA: Sage, pp. 331.

- Weiss (1998): Evaluation: Methods for Studying Programs and Policies. Upper Saddle River, N.J.: Prentice Hall, pp. 244.

Historic reference and specialty subjects:

- Adler; Posner (2006): New Foundations of Cost-Benefit Analysis. Cambridge: Harvard University Press.

- Kirkpatrick; Weiss (1996): Cost-Benefit Analysis and Project Appraisal in Developing Countries. Cheltenham, UK; Brookfield, Vt., US: E. Elgar.

- Chowdhury; Kirkpatrick (1994): Development Policy and Planning : an Introduction to Models and Techniques. London ; New York: Routledge, chapter 7.

- Little; Mirrlees (1994): The Costs and Benefits of Analysis: Project Appraisal and Planning Twenty Years on. Cost-Benefit Analysis. P. R. G. Layard and S. Glaister. Cambridge: Cambridge University Press, pp. 199.

- Squire; Tak (1975): Economic Analysis of Projects. A World Bank Research Publication. Baltimore: Published for the World Bank by Johns Hopkins University Press.

- Little; Mirrlees (1974): Project Appraisal and Planning for Developing Countries. New York: Basic Books.

- UNIDO (1972): Guidelines for Project Evaluation. New York: UNIDO.

- Meimberg (1971): Voraussetzungen einer globalen Entwicklungspolitik und Beiträge zur Kosten- und Nutzenanalyse. Berlin: Dunker&Humblot.

4.3.3. The Effects Method (Méthode des Effets)

The Effects Method is a similar but considerably less well-known efficiency analysis method than Cost-Benefit Analysis (CBA). After some controversy about assumptions made in its early version, nowadays' differences between CBA and the Effects Method lie mainly in the analysis process that, in the case of the Effects Method, involves stakeholders more and in different ways and explicitly and separately treats scenarios with and without intervention. Our assessment is summarised in table 4j and explained in more detail in the subsequent text.

Table 4j. Analytic power and analysis requirements for the Effects Method

Effects Method: Analytic Power				
What is the method's analysis level?	Level 2 (assesses the efficiency of an aid intervention in a way that it can be compared with alternatives or benchmarks)			
To what degree is the method clearly and unambiguously defined?	Some guidance exists			
How independent is the analysis result from the choice of a specific evaluator under *ceteris paribus* conditions?	Results obtained by different evaluators are expected to be very similar (if key assumptions are predefined) or may vary somewhat (if key assumptions are made by evaluator)			
How participatory is the method?	Stakeholder input is restricted to data gathering along established analysis criteria			
Effects Method: Analysis Requirements				
		Qualitative	Numerical	Financial	Monetarisable
Data	Input-level (minimum requirement)				X
	Output-level (minimum requirement)				
	Outcome-level (minimum requirement)				X
Time	Overall analysis time needed for evaluator	Several weeks to many weeks			
	Overall time requirements for stakeholders	More than a day per stakeholder involved in the assessment			
Skills	Special skills needed	Advanced economic analysis			

Description. The Effects Method is an economic appraisal method developed in France and adapted in 1976 by Marc Chervel and Charles Prou to aid projects. It is almost exclusively documented in vintage francophone literature as *Méthode des Effets.*

In absence of up-to date literature on the subject, we have conducted a number of interviews with French evaluation professionals in order to update ourselves about this method.[117]

Overall, the Effects Method shares many similarities with Cost-Benefit Analysis (CBA) and most of the comments made regarding CBA apply to the Effects Method as well (and are not repeated here).

The Effects Method has been applied to aid projects as alternative to CBA, for example in several francophone developing countries. Today, its usage as project appraisal method has faded. On the international level, the method is little-known. Only 5 of 24 representatives to the OECD DAC Network on Development Evaluation that were surveyed for this study knew the Effects Method at all and only 2 of 24 had ever applied it.

Nevertheless, the Effects Method has continued to be cited as alternative and complement to CBA. In 1997, a guide to economic analysis of technical assistance projects by the Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ)[118] lists the Effects Method as alternative to CBA, and EuropeAid's project evaluation manual[119] incorporates methodology from both the Effects Method and CBA. Within the Agence Française de Développement (AFD), the Effects Method is used as tool for highlighting specific aspects of development projects or sector-wide activities and is also taught for pedagogy.

The Effects Method provides essentially the same results as a CBA, i.e. the direct and indirect benefits and costs generated by a project over time from which net present benefits are calculated.

The analysis approach is however somewhat different. The economic scenarios with and without the intervention are explicitly and separately considered and assessed in a participatory process together with officials of the aid recipient country. This is done in three steps:

- Step 1: The scenario without intervention is characterised.

- Step 2: The scenario with intervention is assessed. This is done by systematically decomposing the direct and indirect value generation through project activities along a production chain. At each stage, the imported and local inputs are identified and the distribution of value-add is analysed, separately for each stakeholder.

- Step 3: The scenario without project is assessed in exactly in the same way. Based on the results obtained, the net benefit of the intervention is calculated. This step is equivalent to computing the benefits and opportunity costs when conducting CBA.

An assumption often used in the Effects Method is that of constant demand with and without intervention. This assumption reflects the original national planning philosophy of the method. In the 1970s, several additional assumptions used in the Effects Method ignited academic controversy,[120] for example the focus on market prices that lead to zero opportu-

---

[117] The results of these interviews and of our desk research regarding the Effects Method are summarised in Bombart; Palenberg (2010): Friend or Foe? The Effects Method and Cost-Benefit Analysis. Munich: Institute for Development Strategy.

[118] GTZ (1993): Anleitung für die ökonomische Beurteilung von Projekten der Technischen Entwicklungszusammenarbeit. Frankfurt am Main: GTZ, pp. 9.

[119] European Commission (1997): ECO-Fin Manual. Brussels: European Commission. Downloaded from http://ec.europa.eu/europeaid/multimedia/publications/publications/manuals-tools/t103_en.htm last visited on 28 May 2010.

[120] Some interviewees also felt that early controversies were additionally fuelled by different underlying paradigms: a planning approach in the case of the Effects Method versus a free market approach in the case of Cost-Benefit Analysis.

nity costs for some types of labour and introduced a bias towards in-country production, again reflecting its original proximity to central planning approaches.[121]

In recent applications, the method has detached itself from specific assumptions and allows for more flexibility regarding assumptions, e.g. allowing for inclusion of shadow prices and general equilibrium demand functions.

Remaining differences between the Effects Method and CBA consist in the analysis process. While CBA is usually conducted by a single evaluator, the Effects Method is an analysis exercise that involves experts and officials from the recipient country in the valuation of costs and benefits. In addition, the Effects Method explicitly traces costs and benefit generation along the relevant value chains which routinely generates information about distributional effects as well as on several macroeconomic indicators. This is considered an advantage of the Effects Method by its proponents since it automatically provides decision-makers with additional relevant information in addition to the total net benefit.

Analytic power. The Effects Method is a level 2 analysis since it assesses an entire intervention against a scenario without intervention. It produces the same principal welfare indicator as CBA, i.e. the net benefits to society. In addition, information about distributional effects and value added along supply and production chains is automatically produced as part of the analysis.

In terms of documentation, the method seems to have evolved somewhat from what is described in the original publications. While some up-to-date guidance in French language may exist, we have not been able to identify a comprehensive description of the method in English literature.

As in the case of Cost-Benefit Analysis, the stability of analysis results for different evaluators depends on the level of detail of pre-defined assumptions.

Although stakeholders are considerably more involved, their involvement remains restricted to data gathering along established analysis criteria.

Analysis requirements. The Effects Method has very similar data requirements to CBA.

A particularity of the Effects Method is, however, its use of public accounting tables that are often referred to as input-output tables. These tables are usually not available anymore in developing countries. This requires the evaluator to reconstruct some of that information during the analysis. We nevertheless estimate the overall work time requirements to be similar to CBA.

With regard to time requirements for stakeholders, the Effects Method is more expensive and requires, in our estimate, more than a day per stakeholder involved in the assessment.

We also estimate skill requirements to be very similar to those required for conducting CBA.

Good practice and suggestions for extension of the basic design. Due to the similarity between the Effects Method and CBA, the good practice recommendations made in the con-

---

[121] See, for example, the objections by Bela Balassa (Balassa (1976): The "Effects Method" of Project Evaluation. Oxford Bulletin of Economics and Statistics 38; Balassa (1976): The "Effects Method" of Project Evaluation once again. Oxford Bulletin of Economics and Statistics 39) and Marc Chervel's response (Chervel (1977): The Rationale of the Effects Method: a Reply to Bela Balassa. Oxford Bulletin of Economics and Statistics 39) as well as the remark by Little and Mirrlees that the Effects Method is erroneous (Little; Mirrlees (1994): The Costs and Benefits of Analysis: Project Appraisal and Planning Twenty Years on. Cost-Benefit Analysis. Layard and Glaister. Cambridge: Cambridge University Press, p. 199).

text of CBA apply also here with the exception of the remark on inclusion of distributional effects into CBA since these are automatically assessed in the Effects Method.

Suggested literature

- Palenberg (2009): Tools and Methods for Evaluating Aid Efficiency: Inception Report. Bonn: BMZ.

- Chervel (2008): Mondialisation des Marchandises et Marchandisation du Monde: Matériau de Pensée Economique pour le Nouveau Siècle. Paris.

- Piveteau (2004): Évaluer les ONG. Paris: Edition Karthala.

- Franck (1996): The Effects Method and Economic Cost-Benefit Analysis: Substitutes or Complements? Cost-Benefit Analysis and Project Appraisal in Developing Countries. C. H. Kirkpatrick and J. Weiss. Cheltenham, UK ; Brookfield, Vt., US: E. Elgar.

- Ministère de la Coopération (1978): The Methodology of Planning, Manual of Economic Evaluation of Projects, The Effects Method. Paris: Ministère de la Coopération (out of print).

- Chervel (1977): The Rationale of the Effects Method: a Reply to Bela Balassa. Oxford Bulletin of Economics and Statistics 39, 1977.

- Balassa (1976): The "Effects Method" of Project Evaluation. Oxford Bulletin of Economics and Statistics 38.

- Balassa (1976): The "Effects Method" of Project Evaluation once again. Oxford Bulletin of Economics and Statistics 39.

4.3.4. Cost-Effectiveness Analysis

Cost-Effectiveness Analysis (CEA) measures the principal outcome of an intervention in its natural units and contrasts it with the intervention's costs. In this way, CEA can also handle interventions with effects that are hard to express in monetary units. At the same time, CEA is however restricted to interventions with a single principal outcome. Our assessment is summarised in table 4k and explained in more detail in the subsequent text.

Table 4k. Analytic power and analysis requirements for Cost-Effectiveness Analysis

Cost-Effectiveness Analysis: Analytic Power	
What is the method's analysis level?	Level 2 (assesses the efficiency of an aid intervention in a way that it can be compared with alternatives or benchmarks)
To what degree is the method clearly and unambiguously defined?	Clear and established analysis procedures exist
How independent is the analysis result from the choice of a specific evaluator under *ceteris paribus* conditions?	Results obtained by different evaluators are expected to be very similar (if key assumptions are predefined) or may vary somewhat (if key assumptions are made by evaluator)
How participatory is the method?	Stakeholder input is restricted to data gathering along established analysis criteria

Cost-Effectiveness Analysis: Analysis Requirements					
		Qualitative	Numerical	Financial	Monetarisable
Data	Input-level (minimum requirement)			X	(X)
	Output-level (minimum requirement)				
	Outcome-level (minimum requirement)		X		
Time	Overall analysis time needed for evaluator	Several days to several weeks			
	Overall time requirements for stakeholders	A few hours or less per stakeholder involved in the assessment			
Skills	Special skills needed	Basic economic and financial analysis (if based on financial cost analysis) or advanced economic analysis (if based on economic cost analysis)			

Description. If it is difficult or controversial to express benefits and costs in monetary terms and conducting a Cost-Benefit Analysis (CBA) may not be the best option. In such cases, or if a leaner analysis is preferred, Cost-Effectiveness Analysis (CEA) may be a good alternative.

CEA treats costs of an intervention either similarly to CBA or applies simpler cost analysis. The principal difference between both methods is that CEA does not express results of an intervention in monetary terms, but rather considers results in their natural units.[122]

> *Consider, as an example, a medical aid intervention with principal outcome of saving ten lives based on costs of ten million US Dollars. Leaving aside time and chance for the moment, CEA would yield a cost per life saved of one million US Dollars.*
>
> *Decision-makers may feel more comfortable with this number than with analysts reporting positive net benefits of 40 million US Dollars (if saving a life is worth five million US Dollars) or negative net benefits of -5 million US Dollars (if saving a life*

---

[122] The most common definition which we also use in this report excludes monetarisation of results altogether. However, some authors allow for monetarisation of some results and compare residual costs (costs minus monetarised results) with non-monetarised results. See, for example, Levin; McEwan (2001): Cost-Effectiveness Analysis: Methods and Applications. Thousand Oaks, Calif.: Sage Publications, pp. 133.

is worth 0.5 million US Dollars), even if both valuations are based on established methods.[123]

Many remarks made with regard to CBA in section 4.3.2 apply to CEA as well. In what follows, we summarise several features that are specific to CEA:

- *Only relative judgements.* Since costs and effects are measured in different units, net benefits cannot be calculated. Therefore, CEA cannot tell us whether, overall, society has gained or lost as a consequence of an intervention. Anthony Boardman, et al, describe this as follows:

  > *CEA measures technical efficiency, not allocative efficiency. It can rank alternative policies in terms of technical efficiency but cannot indicate whether something is worth doing.[124]*

- *Choosing the right measure for effects.* Since benefits are measured in their natural units, different types of benefits are incommensurable, i.e. not expressed by the same units, and cannot be aggregated into a single, overall benefit. This makes comparison of interventions with multiple outcomes difficult and effectively restricts application of CEA to interventions with one similar principal outcome each.

  One workaround is possible if benefits of all but one outcome can be quantified and monetarised as in CBA. Then, the residual costs, i.e. the costs minus the monetarised benefits, can be compared across interventions to the outcome that has not been monetarised. However, the results of these considerations may be difficult to interpret and it may be recommendable to simply conduct a full CBA instead.

  When choosing a measure for quantifying the principal effects of an intervention, care must be taken that this measure is actually related to benefits in the analysis perspective that is chosen. For example, measures relating to transfers of welfare within a society are irrelevant when taking the perspective of the entire society.

  In addition, effects must be measured against a reference scenario, i.e. compared to the effects without an intervention. It should also be clearly stated whether average or marginal effects are considered.

- *Cost-effectiveness ratios.* Usually, CEA results are expressed as cost-effectiveness ratios, i.e. the cost per unit of effect. Sometimes, the inverse ratio is used as well, i.e. units of effect per cost. Somewhat confusingly, both ratios usually go by the same name.[125]

  When comparing cost-effectiveness ratios, attention must be paid to whether average or marginal costs and effects or if incremental costs and effects between different interventions are used.

  Cost-effectiveness ratios have some similarities to unit costs that were described in section 4.2.1. However, while unit costs are based on output-level results, CEA needs

---

[123] For different methods for valuating saving lives and the example numbers used here in 1988 US Dollars see, for example, Gramlich (1990): A Guide to Benefit-Cost Analysis. Englewood Cliffs, N.J.: Prentice-Hall, pp 67.

[124] Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, p. 483.

[125] See, for example, Levin; McEwan (2001): Cost-Effectiveness Analysis: Methods and Applications. Thousand Oaks, Calif.: Sage Publications, pp. 133.

to consider results that reflect the main benefits of an intervention. This leads to effects measures on the outcome and impact level.

When using cost-effectiveness ratios for the selection of interventions, most remarks regarding benefit-cost ratios in section 4.3.2 apply here as well. Since interventions may not be scalable or divisible, selection of projects with the best cost-effectiveness ratio may not maximise overall benefits in the presence of constraints. In addition, ambiguity in the definition of costs will translate into cost-effectiveness ratios. In contrast to CBA, where these issues could be avoided by relying on net benefits, no simple solution exists in the case of CEA.

- *Scale issues, cost minimisation and yield maximisation.* Until now, we have implicitly assumed that benefits are proportional to the size of the effect that is measured by CEA, i.e. that twice the effect produces twice the benefit.

This is, trivially, the case, if benefits and effects are the same. If, for example, the principal objective of a health intervention is to save lives, the same measure can be used in CEA and we have no problem.

If, instead, CEA uses an intermediate measure, this simple correlation is usually lost as illustrated by the following example.

> *Consider an educational intervention that is aimed at improving the career opportunities of participants. Obviously, this quantity is difficult to measure and a more accessible intermediate effects measure is used: the math test scores increase for participants. Let us assume that without the intervention, participants score 50 out of 100 points on average.*

> *Now consider interventions A through E with the following features:*

Intervention	Cost (in Euro per child)	Math test score after intervention	Math score increase	Cost per percentage point of test score increase
"Do nothing"	0	50	-	-
A	100	55	5	20
B	100	60	10	10
C	200	70	20	10
D	300	70	20	15
E	800	90	40	20

> *Two principal observations can be made:*

>> - *Some interventions have lower efficiency than others. For example, intervention B is preferable to intervention A since, with the same costs, it leads to a higher test score increase. In technical terms, intervention B dominates intervention A. Similarly, Intervention D is dominated by intervention C since both produce the same effect but C costs less.*

>> *These special cases are exactly what is described as yield maximisation and cost minimisation in section 2.2.5.*

>> - *Apart from special cases of dominance, we cannot compare interventions. For example, we cannot decide whether intervention*

B, C or E is preferable. This is due to the fact that we have not quantified how benefits relate to test score increases. This relation is not simple, since benefits connected to increasing average test scores by one percentage point is likely to depend also on the test score itself.

Instead, in the analysis of dominance above we have just assumed that higher are better than lower test scores, without saying by how much.

In both observations, the cost-effectiveness ratios do not play a role and can be misleading.

For example, the fact that the ratios for interventions B and C are identical does not imply that benefits and costs relate in the same way to each other for both options. If benefits generated by a test score increase from 50 to 60 were greater than those generated by a test score increase from 60 to 70, intervention B would be preferable. In the opposite case, intervention C would be preferable.

Generally, whenever effects measures are used that are not proportional to benefits, results of CEAs need to be interpreted with care.[126] In such cases, cost-effectiveness ratios can be misleading and only alternatives that dominate others, either by producing more effect with the same costs (yield maximisation) or by producing the same effect with less cost (cost minimisation), can be identified without making further assumptions.

Some experts interviewed restricted CEA altogether to those special cases. We feel that this is not useful since it excludes applications of CEA when the complications described above are not relevant.

In some cases, CEA is also, erroneously, understood as minimising costs or maximising yield without consideration of the respective other side of the equation as summarised by Diana Fuguitt and Shanton Wilcox:[127]

Cost-effectiveness analysis "compares the costs and effects" of a policy alternative. This is emphasized because mistakes are occasionally made. Applications in the literature sometimes misinterpret cost-effectiveness analysis as identifying the policy alternative that minimizes costs (with no regard to the policy's effectiveness) or that maximizes effectiveness (with no regard to costs).

- *Discounting of effects.* When conducting CEA, costs occurring in different years are discounted just as in CBA. However, experts are less sure about whether or not, and with what discount rate, effects should be discounted. Henry Levin and Patrick McEwan conclude that the evidence for discounting effects is compelling but that experts argue about whether to use the same discount rate used for cost-discounting or a

---

[126] For a more detailed account on non-ratio scale issues in Cost-Benefit Analysis see, for example, Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, pp. 464.

[127] Fuguitt; Wilcox (1999): Cost-Benefit Analysis for Public Sector Decision Makers. Westport, Conn.: Quorum, p. 276.

lower rate. While discounting of effects in health is common, CEA of education projects usually do not apply discounting.[128]

Analytic power. In our assessment, Cost Effectiveness Analysis is a level 2 analysis, i.e. it is capable of comparing entire aid interventions with each other. This obviously implies that the outcomes effects measured translate into impact in the same fashion for all interventions that are investigated.

However, several experts felt that CEA represented a level 1 rather than a level 2 analysis since it essentially measured the production efficiency of outcomes but would otherwise be very similar to the Benchmarking of Unit Costs approach described in section 4.2.1.

Compared to CBA, CEA has the advantage that it can be applied without monetarising results. At the same time, it has the disadvantage of being able to only handle one principal outcome and is therefore not applicable to interventions with several different important outcomes.

CEA is well but somewhat less documented than CBA which may be due to the fact that it partly builds on CBA methodology. Many textbooks explain CBA in detail and describe CEA somewhat shorter as an alternative afterwards.

The analysis itself necessitates fewer assumptions than CBA since outcomes do not have to be expressed in monetary units. If the outcome measure to be used is predetermined, different evaluators are likely to come to very similar results. If instead, specific assumptions within CEA are made by the evaluator, conclusions may differ somewhat.

In many cases, stakeholder input will be included in one or the other way into the analysis along predefined categories.

Analysis requirements. Cost-Effectiveness Analysis is based on monetary input-level data and numerical data on outcomes.

The work time needed for conducting CEA is likely to be less than for CBA since effects do not need to be monetarised. We estimate the total work time for an evaluator to range from several days (if cost and effects data is easily available) to several weeks (if this information needs to be gathered as part of the analysis).

We estimate time requirements for stakeholder to be modest and to remain below a couple of hours per stakeholder involved in the assessment.

The skills needed for conducting CEA depend on what type of cost analysis is conducted. For fairly simple cost analysis, only basic economic and financial skills are required. If an economic cost analysis is conducted, advanced economic analysis skills are needed.

Good practice and suggestions for extension of the basic design. In addition to remarks made for CBA and CEA, we consider the following two points to be helpful.

- A sensitivity analysis may be useful to demonstrate how CEA results depend on underlying assumptions. However, if cost-effectiveness ratios are considered, simple forms of sensitivity analysis may be hard to implement because cost-effectiveness ratios do not have a linear functional relationship to input variables. It may, for example,

---

[128] Levin; McEwan (2001): Cost-Effectiveness Analysis: Methods and Applications. Thousand Oaks, Calif.: Sage Publications., pp. 129.

be very difficult to identify the best and worst case scenarios. If possible, more refined techniques such as Monte-Carlo simulation should be applied.[129]

- If some costs are difficult to assess in CEA, the same costs may be omitted for all interventions. In turn, the numerical values of cost-benefit ratios lose their significance but their ranking remains intact.

Suggested literature

- Levin; McEwan (2001): Cost-Effectiveness Analysis: Methods and Applications. Thousand Oaks, Calif.: Sage Publications.

- Kee (2004): Cost-Effectiveness and Cost-Benefit Analysis. Handbook of Practical Program Evaluation. J. S. Wholey, H. P. Hatry and K. E. Newcomer. San Francisco: Jossey-Bass, pp. 506.

- Rossi; Lipsey; Freeman (2004): Evaluation : a Systematic Approach. Thousand Oaks, CA: Sage, pp. 362.

- Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, pp. 463.

- Eureval-3CE; DG Budget (2006): Study on the Use of Cost-Effectiveness Analysis in EC's Evaluations. Lyon: Centre for European Expertise and Evaluation.

4.3.5. Cost-Utility Analysis

Cost-Utility Analysis (CUA) expresses all outcomes of an intervention by a common utility measure that is then contrasted with the intervention's costs. This makes CUA a widely applicable efficiency analysis method since it does not require monetarisation of benefits as CBA, nor is it restricted to interventions with a single principal outcome as CEA. Our assessment is summarised in table 4l and explained in more detail in the subsequent text.

---

[129] For a more detailed description of ratio problems in sensitivity analysis see, for example, Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, pp. 470.

Table 4l. Analytic power and analysis requirements for Cost-Utility Analysis

Cost-Utility Analysis: Analytic Power	
What is the method's analysis level?	Level 2 (assesses the efficiency of an aid intervention in a way that it can be compared with alternatives or benchmarks)
To what degree is the method clearly and unambiguously defined?	Clear and established analysis procedures exist
How independent is the analysis result from the choice of a specific evaluator under *ceteris paribus* conditions?	Results obtained by different evaluators are expected to be very similar (if key assumptions are predefined) or may vary somewhat (if key assumptions are made by evaluator)
How participatory is the method?	Stakeholder input is restricted to data gathering along established analysis criteria

Cost-Utility Analysis: Analysis Requirements					
		Qualitative	Numerical	Financial	Monetarisable
Data	Input-level (minimum requirement)			X	(X)
	Output-level (minimum requirement)				
	Outcome-level (minimum requirement)		X		
Time	Overall analysis time needed for evaluator	Several weeks to many weeks			
	Overall time requirements for stakeholders	A few hours or less per stakeholder involved in the assessment (if utility standards are used) More than a day per stakeholder involved in the assessment (if relative utilities are measured as part of the analysis)			
Skills	Special skills needed	Basic economic and financial analysis (if based on financial cost analysis) or advanced economic analysis (if based on economic cost analysis) Decision theory and utility theory			

Description. Cost-Utility Analysis (CUA) describes all results of an intervention by their utility, a quantity reflecting the preferences associated with these results. Utilities for different results are combined into a measure for total utility that is then related to the costs of the intervention.

CUA can therefore be understood as a generalisation of Cost-Effectiveness Analysis (CEA). It removes the restriction to one principal outcome inherent in most CEA approaches by measuring all results in the same unit, i.e. by their utility.

Compared to Cost-Benefit Analysis (CBA), CUA does not rely on the willingness to pay concept and does not monetarise results. This may be advantageous if expressing results in monetary terms is difficult or controversial.

The following example illustrates these two considerations:

> *CUA is often applied to interventions that impact both the length and the quality of human lives. Imagine one intervention saving ten lives and decreasing morbidity (i.e. the degree to which a health condition affects a person) in 100 people and another one saving five lives but leading to the same morbidity decrease in 200 people.*

> *For such interventions, CBA would require to attribute a monetarised value to an extra year of life and to a year of life in better health. Both valuations are likely to vary strongly with underlying assumptions and people might therefore reject them*

(or because of fundamental ethical objections against putting a monetary value on life at all).

Conducting CEA would not yield workable results either. We cannot say which intervention produces more benefits because the measures "lives saved" and "improved health" are incommensurable.

CUA, instead, measures both the quantity and the quality of life years with one common measure that is called Quality-Adjusted Life Years (QALYs, sometimes also referred to as Disease-Adjusted Life Years, DALYs).[130] One QALY is equivalent to one year lived in perfect health. Life years lived in less than perfect health receive a QALY score between 0 and 1 and a year not lived (death) receives a QALY score of 0.

In this way, the results of both interventions can be summarised by their total QALY scores. Let us assume that the morbidity decreases in both interventions lead to QALY score increases of 0.1 per person and that the people whose lives are saved are in perfect health. Overall, the first intervention will produce 20 QALYs and the second intervention will produce 25 QALYs for each year these effects hold. If intervention costs are identical, CUA would prefer the second intervention.

CUA results are often expressed as cost-utility ratios, for example as cost per QALY saved (for the above health intervention example) or more generally, as cost per increase of a utility score.

CUA shares some similarities to the MADM approaches presented in section 4.3.1 but differs on three accounts. First, in CUA, costs are treated explicitly and are not included through their own utility function. Second, CUA utility estimates are not based on a single decision-maker's preferences but represent group preferences or standards. Third, CUA builds on numerical outcome data, each unit of which is then assigned a utility value.

Practical application of CUA is concentrated on health-related interventions in which the general CUA concept is often treated synonymously with the standardised QALY concept. Some authors report application to educational interventions, albeit outside of the field of international aid.[131] CUA is, however, cited alongside CBA and CEA in aid evaluation guidelines[132] and in many evaluation textbooks.[133]

The actual measure for utility in CUA is somewhat arbitrary since its absolute value does not matter. In the above medical example, the choice to introduce QALYs that equal one year lived in perfect health is a matter of convenience that helps conceptualising utility. We could also have used any of several Intensive Care Unit morbidity scores, or simply worked with an abstract utility score. In fact, CUA in fields where this approach is less standardised work

---

[130] The concept of DALYs is, among other, applied by the World Health Organization to describe worldwide effects of premature death and disease. See, for example, Mathers; Fat; Boerma (2008): The Global Burden of Disease: 2004 Update. Geneva, Switzerland: World Health Organization. See box on page 3 for WHO's definition of DALYs.

[131] For an application of Cost-Utility Analysis to the assessment of interventions in education see Levin; McEwan (2001): Cost-Effectiveness Analysis: Methods and Applications. Thousand Oaks, Calif.: Sage Publications, pp. 208 and references therein.

[132] See, for example, Castillo (1998): Economic Analysis of Social Investment Fund Projects. Washington, D.C.: World Bank.

[133] See literature at the end of this section.

with abstract utility scores. What matters are only relative values of utilities of different results with respect to each other. In the above example, we somewhat nonchalantly assumed that the quality of life improvements of both interventions, per year, were 10 times less important than extending a life in perfect health for a year.

These relative valuations represent both the cornerstone and the most criticised assumption of CUA.

Several methods for the estimation or measurement of relative utilities are available.[134]

- *Direct methods.* Participants, experts, or other stakeholders are asked to directly rate different outcomes on a scale from 0 to 1 which is then used as cardinal utility scale in CUA.

- *Standard gamble.* The mental lottery described earlier[135] is used to assign utility values to different outcomes by comparing them to a (hypothetical) gamble with best and worst outcomes.

  A somewhat related method is the time trade-off method, in which participants are asked to compare utilities for different outcomes through trade-off considerations, for example by determining how many "healthy life years" are considered equivalent to ten years of life with a specific less-than-perfect health status.

- *Standardised indexes.* Mostly developed in health policy interventions, classification schemes with precise descriptions of different states along a sometimes large number of health dimensions are used to assign utility values.

Critics consider relative utility valuations with these methods to be difficult or controversial because of the following reasons:[136]

- All of these methods aim at determining group utilities and therefore require interpersonal comparisons of preferences; and

- Empirical evidence shows that, in some cases investigated[137] results obtained by different methods for assessing relative utilities differ substantially which casts doubts on the reliability of all methods described above.

In addition, scale issues discussed in the context of CEA in section 4.3.4 apply to CUA as well. If the utility measure chosen does not scale with benefits (as defined by an intervention's principal objectives), cost-utility ratios become meaningless and CUA results can only be used to discard those alternatives that are dominated by others, i.e. that have less utility but the same cost or the same utility and more cost.

Similarly, discounting utility and the inclusion of chance into these approaches needs to be considered as in CBA and CEA.

---

[134] See, for example, Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, pp. 476, or Levin; McEwan (2001): Cost-Effectiveness Analysis: Methods and Applications. Thousand Oaks, Calif.: Sage Publications, pp. 194.

[135] See quantification of individual preferences in section 2.2.2 and example in the description of the scientific decision analysis MADM method in section 4.3.1.

[136] For a critical analysis of the standard gamble approach and aggregate utilities see, for example, Hildred; Beauvais (1995): An Instrumentalist Critique of Cost-Utility Analysis. Journal of Economic Issues.

[137] See, for example, Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, p. 481 and references therein.

Analytic power. Cost-Utility Analysis is a powerful level 2 analysis approach. It combines the analytic rigour of a CBA with the advantage of expressing benefits in units that may be more intuitive than money and avoids conflicts associated with controversial monetarisations.

Especially for health-related interventions, CUA is well-defined in literature. For other applications, the general approach is described in several textbooks.

Similar to CBA and CEA, different evaluators are likely to come to similar conclusions if key assumptions, such as the choice of utility measures, are predefined. Instead, if these assumptions are made as part of the analysis process, results may vary.

Stakeholder feedback is needed when establishing the standards by which relative utilities of different outcomes are valued. In principle, this calibration can be repeated for each time CUA is applied but in practice, whenever possible, established standards are used in order to increase comparability and analysis efficiency.

Apart from this type of stakeholder involvement, CUA gathers data from stakeholders for the measurement of effect sizes, i.e. along predefined categories.

Analysis requirements. Data needed on the input-level for CUA is financial or monetarisable, depending on what cost analysis method is chosen. On the outcome-level, numerical information is usually required.

CUA work time requirements exceed those of CEA since relative valuations of utilities may need to be established and several outcomes are tracked. We estimate that overall work time requirements are similar to those for CBA.

Work time requirements for stakeholders are similarly modest as those for CBA if CUA is based on established standards for relative utilities. If, however, relative utilities are measured as part of the analysis, we estimate stakeholder time investments to exceed a day per stakeholder that is involved.

CUA is one of the most demanding analysis techniques presented in this report. In addition to basic or advanced economic analysis skills, a good understanding and practical experience of utility theory is required.

Good practice and suggestions for extension of the basic design. In addition to the remarks made for CBA, CEA and CUA above, we consider the following remark useful:

CUA produces cost-utility ratios that can be compared to similar ratios obtained for different interventions, for example by compiling tables comparing costs required for saving lives (or saving QALYs) in different interventions.[138] However, care should be taken when making such comparisons since it is easy to compare apples with pears:

- Different CUAs may measure utilities and costs differently or make otherwise differing assumptions;

- Cost-utility ratios suggest scalability of interventions which may not be possible; and

- Different interventions may have different goals that are not visible in the utility measure.

Suggested literature. We have not been able to identify a textbook entirely dedicated to CUA. We consider the textbook by Ralph Keeney and Howard Raiffa (cited below) as the most profound synthesis on utility theory but it does neither cover the separate treatment of

---

[138] For a discussion on issues with league tables see, for example, ibid., pp. 482, and literature within.

cost nor the assessment of group utilities.[139] Other textbooks cited here contain chapters on CUA.

- Keeney; Raiffa (1993): Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge England; New York, NY, USA: Cambridge University Press.

- Boardman (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, pp. 474.

- Weirich (2001): Decision Space: Multidimensional Utility Analysis. Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge, UK; New York: Cambridge University Press.

- Levin; McEwan (2001): Cost-Effectiveness Analysis: Methods and Applications. Thousand Oaks, Calif.: Sage Publications , pp. 189.

5. Overview of Methods and General Observations

In the previous chapter, we have presented different methods for assessing aid efficiency. These methods differed in many ways, for example in the results they provide, in the reliability of these results, in the degree of stakeholder involvement, and in data, time and skill requirements.

In order to put individual methods into context, we synthesise an overview of these methods in the next section (section 5.1) and discuss several general observations.

These observations provide the basis for section 5.2, the last section of this report, in which we return to the principal motivation for our research: the gap between expectations and delivery of efficiency analysis. We offer four recommendations on how to close this gap.

5.1. Overview of Identified Methods

Table 5a shows all 15 methods that have been identified and described in this report.[140] Apart from categorising methods according to their type, we sorted them according to our perception of how well-known they were to the evaluation experts we have interviewed and surveyed.

---

[139] Instead, costs are assessed by their utility and all utility assessments are made from a one-person, decision-maker perspective which, however, doesn't exclude taking into account other peoples' utility assessments by the decision-maker.

[140] We list some methods separately that were discussed in parallel in the last chapter.

Table 5a. Degree to which efficiency analysis methods are known by evaluation experts

Degree to which method is known	Level 2 methods	Level 1 methods	Descriptive methods
Well-known methods	Cost-Benefit Analysis (CBA)	Benchmarking of unit costs	Expert judgement
Somewhat less well-known methods	Cost-Effectiveness Analysis (CEA)	Follow the Money	
		Financial analysis	
		Stakeholder-driven approaches	
		Benchmarking of partial efficiency indicators other than unit costs	
Methods unknown to a substantial fraction of evaluation experts	Multi-Attribute Decision-Making (MADM): Intuitive scoring models	Comparative ratings by stakeholders: Comparative rating of effectiveness and cost analysis	Specific evaluation questions on efficiency
	Multi-Attribute Decision-Making (MADM): Scientific decision analysis		
	Effects Method	Comparative ratings by stakeholders: Comparative rating of efficiency	
	Cost-Utility Analysis (CUA)		

When considering table 5a, it should be kept in mind that we have included five level 1 methods and one descriptive method that are little or not at all described in literature and applied in appraisal and evaluation reports.[141] We have named and described these methods based on observed practice, suggestions by experts and our own common sense. It is therefore not surprising that some experts are not aware of them.

It nevertheless remains surprising that even well-documented level 2 methodology is sometimes little-known. For example, of 24 representatives to the OECD DAC Network on Development Evaluation, when prompted, 33 per cent didn't know Cost-Effectiveness Analysis, 48 per cent didn't know Cost-Utility Analysis and 79 per cent didn't know the Effects Method.[142] This survey did not cover Methods for Multi-Attribute Decision-Making (MADM) that were however equally unknown to many experts we interviewed. In contrast, 92 per cent of respondents knew Cost-Benefit Analysis. While not new to aid evaluation ourselves, this observation also applies to us. Before beginning efficiency-related research, we were unaware of the Effects Method and scientific, utility-theory based version of the MADM approach. We will return to these observations when making some general remarks in the next section.

---

[141] These methods cover most level 1 methods (*Follow the Money, stakeholder-driven approaches, benchmarking of partial efficiency indicators other than unit costs,* and both types of *comparative ratings by stakeholders*) and the descriptive method *analysis based on specific evaluation questions on efficiency*.

[142] For Cost-Utility Analysis there were only 23 respondents.

Table 5b. Degree to which efficiency analysis methods are applied in appraisal and evaluation reports

Degree to which method applied in appraisal and evaluation reports	Level 2 methods	Level 1 methods	Descriptive methods
Often, at least for certain types of aid interventions	Cost-Benefit Analysis (CBA)	Benchmarking of unit costs	Expert judgement
		Financial analysis	
Occasionally	Cost-Effectiveness Analysis (CEA)	Follow the Money	Specific evaluation questions on efficiency
	Effects Method	Benchmarking of partial efficiency indicators other than unit costs	
Seldom or never	Multi-Attribute Decision-Making (MADM): Intuitive scoring models	Stakeholder-driven approaches	
	Multi-Attribute Decision-Making (MADM): Scientific decision analysis	Comparative Ratings by stakeholders: Comparative rating of effectiveness and cost analysis	
	Cost-Utility Analysis (CUA)	Comparative ratings by stakeholders: Comparative rating of efficiency	

As described in section 1.2, we have also reviewed a number of appraisal and evaluation reports for applied efficiency-analysis methodology. This was done opportunistically with the aim of identifying relevant applied methodology, for example by searching for efficiency and methods-related keywords. The sample of about 100 reports that were screened is therefore not representative. Nevertheless, we have developed a rough impression on the frequency with which different methodologies are applied in practice that has been corroborated by our expert interviews and references to applications in textbooks and other publications. In table 5b we summarise these observations.

By and large, this table is similar to table 5a. This is not surprising since both observations are likely to be correlated: little-known methods are applied less often and less frequently applied methods are, in turn, less documented and therefore less well-known.

Based on both observations, we however conclude that considerably more efficiency analysis methods are available than regularly applied. Since evaluation experts often do not know these methods, they may represent an untapped methodological potential.

In the remainder of this section, we will explore this untapped potential further. For each analysis level, we assess analytic power and summarise data, time and skills requirements.

5.1.1. Level 2 Methods

In section 3.1.1, we have defined level 2 analysis as follows:

> *Level 2 analysis is capable of assessing the efficiency of an aid intervention in a way that it can be compared with alternatives or benchmarks.*

Table 5c gives an overview over the applicability of the six level 2 methods discussed in this report. We assess applicability by looking at what data types for measuring results are per-

mitted for each method. In contrast to the way we assessed data requirements of single methods in the previous chapter, we do not restrict this assessment to minimum data requirements but, instead, describe what type of data the methods can be based on.

Table 5c. Applicability of level 2 methods for different data types

Analysis method	Results can be monetarised	Results can only be expressed numerically		(Some) important results can only be expressed qualitatively
		One principal outcome	Several principal outcomes	
Cost-Benefit Analysis (CBA)	X			
Effects Method	X			
Cost-Effectiveness Analysis (CEA)	X	X		
Cost-Utility Analysis (CUA)	X	X	X	
Multi-Attribute Decision-Making (MADM): Intuitive scoring models	X	X	X	X
Multi-Attribute Decision-Making (MADM): Scientific decision analysis	X	X	X	X

If organised in this way, Cost-Benefit Analysis and the Effects Method are most restricted in terms of their data requirements. Cost-Utility Analysis is more generally applicable since it requires numerical but not financial information on results. A somewhat special case is Cost-Effectiveness Analysis that does not require monetarisation either but is restricted to interventions with one principal outcome. Finally, MADM approaches can be applied to all types of interventions. In addition, these methods do not necessarily require an economic or financial cost analysis.

However, the flexible data acceptance of MADM methods comes at the price of strong subjectivity of the analysis results as shown in table 5d. This table summarises our subjectivity assessments of individual methods in chapter 4. For the first four methods, our assessment depends on whether key analysis assumptions[143] are predefined (crosses in the left column) or are made by the evaluator (crosses in the middle column).

---

[143] For example the choice of the discount rate, the time horizon to consider, choice of shadow prices, choice of techniques for the valuation of intangible effects, and so forth.

Table 5d. Degree of subjectivity of level 2 analysis results

Analysis method	All other things remaining equal (*ceteris paribus*) ...		
	... results obtained by different evaluators are expected to be very similar	... results obtained by different evaluators may vary somewhat	... different evaluators (or decision-makers) can come to entirely different conclusions
Cost-Benefit Analysis (CBA)	X	X	
Effects Method	X	X	
Cost-Effectiveness Analysis (CEA)	X	X	
Cost-Utility Analysis (CUA)	X	X	
MADM: Intuitive scoring models			X
MADM: Scientific decision analysis			X

This subjectivity is due to the fact that MADM approaches base their valuations on the preference structure of a single individual, the decision-maker. This decision-maker assigns *utilities* to different scenarios. The values of these utilities may or may not mirror the results of less subjective analysis. An economist decision-maker might be strongly influenced by the results of Cost-Benefit Analyses when assigning utilities. In contrast, a person critical to economical analysis might be more influenced by other information.

This stands in stark contrast to other level 2 methods that valuate costs and benefits according to established guidelines and standards. For these methods, subjectivity enters only by means of the assumptions made that are inspired by professional know-how and not primarily by personal preferences of the evaluator or a decision-maker.

Finally, table 5e summarises the work time and skills required for conducting these analyses.

Cost-Effectiveness Analysis and intuitive scoring models are the lightest methods with regard to work-time requirements. All other methods require several to many weeks of analysis.

In terms of skills, all but the MADM approaches require basic or advanced economic analysis skills. Cost analyses in Cost-Effectiveness Analysis and Cost-Utility Analysis are sometimes simplified by only considering financial costs which reduces the scope of the result but also the demands on the evaluator's economic analysis skills.

In addition, Cost-Utility Analysis and scientific decision analysis require the evaluators to be acquainted with utility and decision theory, which represent skills not usually found in international development agencies or consultant evaluators.

Table 5e. Work time and skill requirements for level 2 methods

Analysis method	Work time requirements	Time requirements for stakeholders	Skills needed that exceed usual evaluation skills
Cost-Benefit Analysis (CBA)	Several to many weeks	A few hours or less per stakeholder involved in the assessment	Advanced economic analysis
Effects Method	Several to many weeks	More than a day per stakeholder involved in the assessment	Advanced economic analysis
Cost-Effectiveness Analysis (CEA)	Several days to several weeks	A few hours or less per stakeholder involved in the assessment	Basic or advanced economic analysis
Cost-Utility Analysis (CUA)	Several to many weeks	A few hours or less per stakeholder involved in the assessment (if utility standards are used) More than a day per stakeholder involved in the assessment (if relative utilities are measured as part of the analysis)	Basic or advanced economic analysis Decision analysis and utility theory
MADM: Intuitive scoring models	A day or less to several days	A few hours or less per stakeholder involved in the assessment	-
MADM: Scientific decision analysis	Many weeks	A few hours or less per stakeholder involved in the assessment	Decision analysis and utility theory

Keeping the above observations on applicability, subjectivity and time and skill requirements of these six level 2 methods in mind, we draw the following conclusions.

- Several alternatives exist to Cost-Benefit and Cost-Effectiveness Analysis that remain applicable if the data requirements of these methods cannot be fulfilled:
  - On the one hand, Cost-Utility Analysis is a viable alternative if multiple numerical outcomes exist that can be assessed according to their relative utility.
  - On the other hand, Multi-Attribute Decision-Making (MADM) methods can be used if some or all costs and benefits cannot easily be quantified. These approaches can be applied as quick assessments, as with intuitive scoring models, or with considerable scientific rigour, as in scientific decision analysis, and can also be used to complement other analysis.
- The Effects Method is closely related to Cost-Benefit Analysis regarding its applicability and analysis results. It has been somewhat neglected by many theorists and practitioners and it might be beneficial to further explore the potential advantages of those aspects in which the Effects Method differs from Cost-Benefit Analysis.

5.1.2. Level 1 Methods

In contrast to level 2 analysis, level 1 analysis does not compare entire aid interventions in terms of their efficiency, but searches for efficiency improvement potential within a given intervention. We have defined level 1 analysis is as follows (see section 3.1.1):

Level 1 analysis is capable of identifying efficiency improvement potential within an aid intervention.

As shown in table 5f, level 1 methods vary strongly in terms of work time requirements, ranging from methods requiring only one or a few days to methods requiring several or many weeks of analysis.

Table 5f. Work-time and skill requirements for level 1 analysis methods

Analysis method	Work time requirements	Time requirements for stakeholders (per stakeholder involved in the assignment)	Skills needed that exceed usual evaluation skills
Benchmarking of unit costs	A day or less to several days	-	-
Benchmarking of partial efficiency indicators other than unit costs	A day or less to several days	-	-
Comparative ratings by stakeholders: Comparative rating of efficiency	Several days	A few hours or less	-
Comparative Ratings by Stakeholders: Comparative rating of effectiveness and cost analysis	Several days	A few hours or less	-
Follow the Money	Several days to several weeks	A few hours or less	-
Financial analysis	Several days to several weeks	A few hours or less	Basic economic and financial analysis
Stakeholder-driven approaches	Several to many weeks	More than a day	Strong qualitative and mixed-method analysis and moderating skills

The first four approaches in this table require only analysis work times of several days or less. For the first two approaches, this is only true if reliable cost data is available to the evaluator. If such data has to be generated as part of an evaluation or appraisal, time requirements for the evaluator (and stakeholders involved in the analysis) are likely to be higher. All four methods can be considered opportunistic approaches that do not exhaustively screen an intervention but rather search for and reveal punctual efficiency improvement potential. This implies that different evaluators may choose to search for efficiency improvement potential in different places, for example by comparing different types of unit costs or asking stakeholders different efficiency-related questions.

Instead, the three remaining approaches follow a more systematic and exhaustive approach. The Follow the Money approach screens all activities, outputs and outcomes that result from an intervention's expenditures. Financial analysis – as defined in this report – examines the entire business case of an investment embedded within an intervention. Stakeholder-driven approaches screen interventions based on what is perceived to be most important by stakeholders.

All level 1 approaches identify partial efficiency improvement potential. It is up to the evaluator to convincingly show that these partial efficiency improvements will effectively translate into overall improvements allocation efficiency, as illustrated by this example:

Consider the case of a training programme. An evaluator might choose to conduct a unit cost benchmarking of the cost per training day per participant. Even if the analysis respects all standards for good benchmarking, e.g. comparing trainings in similar locations, on similar topics, of similar length and participant number

and based on similar training methods and so forth, the unit cost benchmarking itself will only be of limited value. If, for example, training costs of an "expensive" training programme are reduced, this will save costs but may also impact training effects. Without further considerations, the evaluator cannot decide whether high training unit costs are justified by strong outcomes or reflect waste of resources. In order to produce meaningful recommendations, the evaluator needs to understand what effect sizes different trainings are likely to trigger. If the benefit caused by a training that is three times as expensive as an average training are likely to be five times as large, the training is efficient. If the expensive training produces only average effects, its higher costs are not justified.

As shown by this example, careful consideration of the limitations of level 1 analysis is required in order to derive meaningful recommendations. Neglect of these additional considerations was the main concerns of experts sceptical to level 1 approaches.

In terms of their minimum data requirements, as summarised in table 5g, level 1 methodology is considerably more flexible than most level 2 methods.

Table 5g. Minimum data requirements for level 1 methods

Analysis method	Type of data required on the ...		
	Input-level	Output-level	Outcome-level
Benchmarking of unit costs	Financial	Numerical	-
Benchmarking of partial efficiency indicators other than unit costs	Numerical	Numerical	-
Follow the Money	Financial	Qualitative	Qualitative
Financial analysis	Financial	-	Financial
Comparative ratings by stakeholders: Comparative rating of efficiency	Qualitative	Qualitative	Qualitative
Comparative Ratings by Stakeholders: Comparative rating of effectiveness and cost analysis	Financial	Qualitative	Qualitative
Stakeholder-driven approaches	Qualitative	Qualitative	Qualitative

These less demanding data requirements translate into increased applicability. It is hard to think of interventions that would not allow any level 1 methodology to be applied.

Somewhat surprisingly, several level 1 methods are little-known and not often applied in practice (see tables 5a and 5b at the beginning of this chapter). For example, both methods based on comparative rating by stakeholders are ill described in literature and little applied in evaluation reports. However, most evaluation experts we have questioned about the merit of such analysis agreed that, while not being comparable to level 2 analysis in terms of analytic power, it would represent useful analysis and usually be preferable to not conducting any analysis at all.

Overall, we concluded that a number of little-used level 1 methods exist that are able to identify efficiency improvement potential.

5.1.3. Descriptive Methods (Level 0 Analysis)

In this report, we have also included two descriptive methods: expert judgement and specific evaluation questions.

The most well-known and, in our estimate, by far the most frequently applied method in describing efficiency of technical assistance interventions is expert judgement. The term expert judgement is, however, widely used. In this report, we have narrowed it to methods in which an evaluator expresses his or her professional opinion on efficiency based on his or her own personal reference frame, experiences and expectations.

We have been struggling with the fact that this specific type of expert judgement is so popular since, in our view, it has little analytic value and is often executed in a way that reduces its credibility:

- The fact that expert judgements assess efficiency relative to the evaluator's own criteria and expectations restricts its usefulness in a very principal way. Different evaluators can come to entirely different conclusions about the same intervention. This also makes comparisons between expert judgements meaningless if they are conducted by different evaluators.

- In addition, many expert judgements on efficiency found in evaluation reports lacked information on the basis for that judgement, for example in terms of data, observations made or underlying rationale. Often, only the result of the judgement was described. When presented in this way, expert judgements loose credibility since it becomes impossible for an outside person to differentiate between judgements based on sound and critical analysis and random guesses.

One possible explanation for the popularity of expert judgements (as defined in this report) may be that they are the easiest way to conduct any assessment of efficiency at all and thereby formally satisfy efficiency-related evaluation policy or accountability requirement. We however consider the degree of insight gained on the basis of such assessments to be very low if underlying rationale is missing.

On the contrary, specific evaluation questions on efficiency, again defined in more narrow terms in this report than when commonly used, largely exclude the evaluator's personal preferences and can produce a reliable snapshot of some efficiency-related indicators. This information may later be used to draw additional insights if the same standardised set of questions is applied to several interventions.

5.2. General Observations and Recommendations

When introducing the topic of this report, we had motivated our research by the observation that a gap exists between what is expected and what is delivered in terms of efficiency analysis.

In this concluding section, we present our observations and recommendations on how to close this gap. Our comments are fourfold:

- First, we offer our observations on how current methodology for efficiency analysis can be applied more widely and more intensely.

- Second, we discuss further developments of methodology.

In our estimate, these two measures have the potential to reduce the gap considerably. In addition, it will be important to specify expectations regarding the results of efficiency analysis more clearly:

- Third, we discuss how the purpose of efficiency analysis should be specified to match what is realistically possible.

- Fourth, we examine the question of when to apply efficiency analysis at all.

5.2.1. Exploit the Potential of Existing Methodology

During our research, we have seen several indications that existing methodology for assessing the efficiency of aid interventions can be applied on more occasions and with higher quality.

For example, even established and well-known methods are sometimes not applied with sufficient quality and with declining frequency.

A recent report[144] of the World Bank Independent Evaluation Group laments the declining frequency with which Cost-Benefit Analysis (CBA) is applied to Bank projects: from about 70 per cent in the 1970s to about 30 per cent in the early 2000s and in recent years.

The underlying reasons can only partly be traced back to a shift from "hard" towards "soft sectors" that are less conducive to Cost-Benefit Analysis. The frequency with which CBA is applied has also declined substantially in sectors with strong CBA tradition.

Similarly, several quality issues in CBA analyses are reported, including fundamental analytical issues such as lack of transparent presentation of costs and benefits, of inclusion of risk, of public goods character of results, of comparison with reference scenarios and infrequent usage of shadow prices and other technical adjustments to capture social benefits and costs.

Overall, the share of projects with acceptable or good economic analysis in appraisal documents has declined from 70 per cent in a 1990 assessment to 54 per cent in an assessment conducted in 2008.

Other methods are frequently applied but suffer quality issues, such as in the case of unit cost benchmarking and expert judgement. Unit cost comparisons without additional consideration of outcome-level effects may produce misleading results and expert judgement lacking transparency, rationale or data can generate only very limited insights.

Still other methods presented in this report are little-known even by evaluation experts and therefore seldom applied. In the previous sections, we have discussed several methods that might be suitable for more frequent application.

Based on methods and good practices presented in this report, there is ample room for applying well-known methodology more often, for applying little-known methodology at all, and for improving the general quality with which methods are applied.

We do not advocate that all analysis that is theoretically possible should also be conducted. At the end of these comments, we will discuss when and how to apply efficiency analysis. We however advocate that the existing methodological potential, both in terms of different methods and in terms of good practice, should be better exploited.

---

[144] See, World Bank Independent Evaluation Group (2010): Cost-Benefit Analysis in World Bank Projects. Washington: World Bank (received from IEG task manager Andrew Warner on 28 May 28 2010).

Obviously, this requires knowledge of available methods as well as the appropriate skills for conducting the assessments. Some methods presented in this report require advanced skills in fields of expertise not typically populated by evaluators. Carol Weiss describes this as follows:[145]

> *It is probably worth saying at the outset that although efficiency analysis is a logical follow-on from evaluation, it is a specialized craft that few evaluators have mastered.*

We recommend such skills for the evaluators conducting the assessments but also for those commissioning the analysis because of the following three reasons:

- First, without such knowledge, impractical or inappropriate methodology may be selected. Even if profound knowledge of some methods is present, this may still not cover all available methodology. We have been, for example, somewhat surprised by how little methodological overlap there was between efficiency analysis commissioned by economists and analysis commissioned by non-economists.

- Second, if those commissioning the evaluation do not possess the necessary efficiency assessment skills, guidance and quality control for those conducting the assignment will be weak. In our view, a lack of skills is one principal reason for unspecific terms of reference that, in turn, may lead to low assessment quality. For example, a report of the European Commission's use of Cost-Effectiveness Analysis comments on the underlying terms of reference for these analyses:[146]

> *The terms of reference tend to express efficiency related demands that are not fully specified and that cover only a part of the necessary items (efficiency of what? efficiency in achieving what? efficiency in comparison to what? efficiency question asked for what purpose).*

- Third, a sound understanding of suitable methods is a prerequisite for selecting expert evaluators with sufficient experiences and skills.

Apart from ensuring knowledge and skills, the following feature of all level 2 and some level 1 analysis methods may require a fundamental change in the way efficiency analysis is planned and conducted.

Most appraisals and evaluations of aid interventions are conducted in a stand-alone way. We refer to these as *vertical* assessments that analyse several evaluation criteria for a single intervention. This is a suitable approach for most evaluation criteria, but not necessarily for efficiency.

Many efficiency analysis methods require the evaluator to explicitly compare different interventions.[147] For level 2 analysis, this is the case whenever efficiency analysis produces non-standardised results, for example if cost-utility or cost-effectiveness ratios are compared, if methods for Multiple-Attribute Decision-Making are applied, or if results of Cost-Benefit

---

[145] Weiss (1998): Evaluation: Methods for Studying Programs and Policies. Upper Saddle River, N.J.: Prentice Hall, pp. 244.

[146] Eureval-3CE; DG Budget (2006): Study on the Use of Cost-Effectiveness Analysis in EC's Evaluations. Lyon: Centre for European Expertise and Evaluation, cited from the executive summary.

[147] All efficiency analysis methods assess costs and benefits against the scenario without intervention. We refer here to the case that efficiency-related information needs to be assessed explicitly for several different interventions.

Analyses with incomplete coverage[148] are compared. For level 1 analysis, this occurs whenever non-standardised benchmarks are used.

For these methods, in order to produce meaningful results, efficiency needs to be assessed across several interventions. We refer to these as *horizontal* assessments.

When planning evaluations involving these methods, provisions for conducting horizontal assessments have to be made. Since evaluators will have to assess several interventions in parallel, it may not be the best option of conducting this horizontal assessment of efficiency as part of an otherwise vertical evaluation. Alternatively, horizontal efficiency assessments can be conducted separately, for example based on basic efficiency-related information collected during various vertical assessments.

For other efficiency analysis methods, horizontal assessments are not necessary. This is, for example, the case if the analysis produces standardised results[149] such as Cost-Benefit or Cost-Utility Analysis that can be compared with existing benchmarks or if limited analysis such as the Follow the Money approach is conducted that does not require comparison.

5.2.2. Further Develop Promising Methodology

Some of the methods presented in this report are considerably less developed than others. In this context, a senior World Bank evaluation expert pointed to the tremendous research efforts that had brought Cost-Benefit Analysis to its present level of sophistication and extrapolated that similar research might bring "softer" approaches to a similar level of analytic power.

While not entirely sharing this level of optimism we feel that the potential of several methods outlined in this report can be developed considerably.

First, a number of simple and common-sense based approaches may benefit from professionalisation and standardisation. For example, both types of comparative stakeholder ratings need to be researched in terms of their reliability and generalisability. For similar types of interventions, it may be useful to establish standardised questionnaires that minimise bias and maximise specificity. Possibly, public databases for specific benchmarking purposes are useful as well. Expert judgement approaches may be combined with specific evaluation questions to reduce ambiguity and subjectivity in what may be similar to trained observer ratings.[150] These are, however, just initial examples. Better ways to improve little-developed methodology or even to add new methods to this report's catalogue can most likely be developed.

Second, highly sophisticated approaches may benefit from rendering them more practicable. For example, Cost-Utility Analysis has been virtually absent in the efficiency analysis of aid interventions that do not have a direct and combined effect on mortality and morbidity and base their analysis on the concept of Quality-Adjusted Life Years (QALYs). Applying Cost-Utility Analysis for other fields requires developing a new set of utility measures and measuring general preferences before being able to start with the analysis itself. Obviously, if (unknown) utility measures exist these should be investigated first. This represents an extra ef-

---

[148] For example if certain costs or benefits equal in all alternatives are omitted.

[149] In order to produce standardised results it is not only necessary to apply the methodology in the same way, but also to base it on comparable, realistic assumptions.

[150] Greiner (2004): Trained Observer Ratings. Handbook of Practical Program Evaluation. Wholey, Hatry and Newcomer. San Francisco: Jossey-Bass, pp. 211.

fort, a first-mover disadvantage and an effective barrier for the development of Cost-Utility Analysis approaches in new fields. Considering that the QALY-concept was not developed for aid evaluation but rather for assessing effects of medical interventions in developed countries, research on other existing applications outside the field of development aid or a comprehensive effort for developing aid-specific utility measures might be useful. Similarly, scientific decision analysis, one of the two Multi-Attribute Decision-Making methods portrayed in this report, may be hard to understand and apply for non-academicians. During our interviews, some experts remembered this type of analysis from their time at university but didn't connect it to assessing aid efficiency. This is unfortunate since the analysis process and the underlying principles are straightforward. Here, a readable synthesis of the analysis process and research into sensible simplifications, both tailored specifically to aid interventions, might be useful.

5.2.3. Specify Realistic Expectations

Efficiency analysis can be a powerful analytical instrument. It can support decision-making by synthesising complex empirical evidence into simple efficiency measures and it can identify efficiency improvement potential in a systematic and reliable way.

As an instrument, efficiency analysis also has natural limitations. These can be fundamental or operational in nature:

- On the one hand, fundamental limitations to efficiency analysis are related to limitations of the paradigms specific models are based on.

  For example, methods for Multi-Attribute Decision-Making value costs and benefits from the perspective of a decision-maker. Even if we assume that this decision-maker has access to sufficient information, it is unclear to what extent this information influences his or her assessment. If exclusively applied from the perspective of the entire society, Cost-Benefit Analysis has a blind eye towards distributional effects (unless societal sub-groups are considered as well) and Cost-Effectiveness Analysis can usually cover only one principal outcome.

- On the other hand, operational limitations are given by the accuracy with which the method is implemented, for example by the quality of assumptions made by the evaluator or the diligence of his or her work.

Under some circumstances, and for some types of aid interventions, these limitations may be so important that a method is not applicable anymore. These natural limitations need to be reflected in expectations of efficiency analysis: the purposes of efficiency analysis need to be stated more clearly. The introduction of levels of efficiency analysis in this report allows for a more specific description of the realistic possibilities of efficiency analysis:

- It is our conviction that, for *project-type interventions*, all analysis levels are principally possible. Based on the methods offered in this report, a sufficiently skilled and resourceful evaluator with sufficient capacity should be able to produce a comprehensive relative valuation of project alternatives (level 2 analysis) and to identify efficiency improvement potential (level 1 analysis).

  This, however, does not imply that the efficiency of projects can generally be compared across the board. Level 2 methods usually compare efficiency only within a restricted group of interventions.

- *Simple programme-type interventions* that are composed of many projects can, in principle, be assessed by the sum of many project-level efficiency assessments. This is easy if programmes are composed of many similar projects and cumbersome if all individual projects are different.

  If, instead, the assessment is conducted directly on the programme-level, i.e. without aggregating project-level inputs and results in a bottom-up way, level 2 analysis, in our opinion, becomes harder to conduct and is not solidly grounded in empirical observation of costs and benefits. Some level 2 methods are able to deal with incomplete empirical information but pay this flexibility with increased subjectivity.

- *Highly aggregated aid modalities.* The trend of increased difficulty in applying level 2 analysis continues for more aggregated aid modalities. Based on current or near-future methodology, we feel that it would be far-fetched to expect solid analysis that can, for example, compare the development efficiency of different budget support interventions with each other. The same would apply to other aggregated bi- or multi-lateral aid modalities. For these types of aid, benefits and costs are difficult to assess empirically. For aggregated aid modalities, results of level 2 analyses are therefore likely to imply unreliable assumptions.

  Methods for Multi-Attribute Decision-Making remain applicable but are equally disconnected from empirically observed costs and benefits. Interestingly, the comparatively simple logic of scoring models that weigh and rate decision-making attributes according to their utility seems to be implicit in some argumentations in favour or against aggregated aid modalities, for example when describing the utility associated with pooling aid or placing management into the hands of recipient countries. We have however been unable to find any explicit application of MADM approaches in this context.

Obviously, our judgements are personal and may not reflect the opinion of others. In our interviews, however, most experts agreed with the general trend summarised above. Many were somewhat less optimistic regarding the possibilities of level 2 efficiency analysis and a few were more optimistic.

If our general estimates hold true, the expectations of policy-makers as reflected in evaluation principles and criteria, guidance documents and terms of references should be clarified and specified and reflect realistic analysis purposes.

5.2.4. Consider Costs and Benefits of Efficiency Analysis

Efficiency analysis itself produces costs and benefits:

- Analysis costs are usually reflect the time for preparing and conducting the analysis, including time requirements for stakeholders, and for disseminating and implementing analysis results.

- Analysis benefits are, ultimately, increased development impact. These can be reached by means of several pathways, for example by providing assistance for selection of more efficient interventions, by directly improving the efficiency of ongoing or planned interventions, by fostering learning through publication and dissemination of appraisal, evaluation or research reports, or by developing required skills through capacity development measures.

It is therefore important to consider whether or not the benefits of efficiency analysis justify the costs. Among different efficiency analysis options, the type of efficiency analysis producing the largest net analysis benefit should be selected. While it may be difficult to estimate concrete analysis benefits, several basic recommendations can made:

- Level 2 efficiency analysis should serve a concrete purpose. As appraisal technique, it typically informs the selection of interventions. As evaluation technique during or after an intervention, it is used to verify and calibrate appraisal assumptions. If, instead, efficiency analysis is done only for the sake of doing it, analysis benefits cannot justify analysis costs. A recent World Bank Independent Evaluation Group report estimated the costs of conducting a Cost-Benefit Appraisal based on the analyst's time to amount to roughly 16,000 US Dollars and estimated potential benefits in terms of avoiding mistakes or selecting a better project to lie between one and ten million US Dollars. The same report, however, also warns:

  > *If the Bank places little weight on cost-benefit results for decisions, staff will have little incentive to invest effort in it. Mandating that staff conduct cost-benefit analysis while giving it little weight in decisions imposes costs with few benefits.[151]*

- If, before conducting any analysis, it is obvious that an intervention ranks among best in class, a level 2 efficiency analysis may add little value since the intervention would have been selected anyway. The same is true for worst-in-class interventions. Peter Rossi, Mark Lipsey and Howard Freeman[152] phrase this as follows:

  > *[...] efficiency analysis may be unnecessary if the efficacy of the program is either very minimal or very high. Conducting an efficiency analysis makes sense primarily when a program is effective but not perfectly so.*

- Similarly, if the decision to implement an intervention has already been taken, it may be a better investment to conduct one or several level 1 assessments.

- Finally, many superficial efficiency assessments based on expert judgements that we rated as having purely descriptive value have remained without clear purpose to us and the related resources might be better invested in other types of efficiency analysis or for entirely different purposes.

  A similar argument applies if required analysis skills are not available which is likely to result in low-quality analysis that may have zero or even negative effects.

Naturally, any decision to not conduct efficiency analysis, or to conduct it in a limited way, should be clearly and transparently documented.

---

[151] See, World Bank Independent Evaluation Group (2010): Cost-Benefit Analysis in World Bank Projects. Washington: World Bank (received from IEG task manager Andrew Warner on 28 May 2010; unpublished at that time) and World Bank Independent Evaluation Group (2009): Annual Review of Development Effectiveness 2009: Achieving Sustainable Development. Washington: World Bank, p. 15.

[152] Rossi; Lipsey; Freeman (2004): Evaluation : a Systematic Approach. Thousand Oaks, CA: Sage, p. 336.

Annex A: Interviews

A total of 44 interviews have been conducted for this study and are listed below, ordered first alphabetically by institution and second by last name. Brief and superficial contacts and information exchanges are not counted as interviews. In addition, a survey has been conducted with 29 representatives to the OECD DAC Network on Development Evaluation that are not listed by name.

Six face-to-face interviews have been conducted by Diane Bombart of the Institute for Development Strategy for research on the Effects Method.[153] All other interviews have been conducted by the author.

First Name	Last Name	Job description	Home institution	Interview modality
Howard	White	Executive Director	3ie	Face to face
Martina	Jacobson	Project Manager	AGEG Consultants eG	Telephone
Carl	Bernadac	Evaluation Officer	Agence Française de Développement (AFD)	Face to face
Jean-Pierre	Lemelle	Evaluation Officer	Agence Française de Développement (AFD)	Face to face
Olivier	Ratheaux	Evaluation Officer	Agence Française de Développement (AFD)	Face to face
Michel	Griffon	Co-director and Researcher	Agence Nationale pour la Recherche (ANR)	Face to face
Henry	Levin	William H Kilpatrick Professor	Columbia University	Telephone
Daniel	Clauss	Results Oriented Monitoring (ROM) coordinator	EuropeAid	Face to face
Catherine	Pravin	Chargée d'Evaluation	EuropeAid	Face to face
Susanne	Wille	Evaluation manager	EuropeAid	Face to face

---

[153] Interviewed experts were Carl Bernadac, Michel Griffon, Philippe Hugon, Michel Le Gall, Jean-Pierre Lemelle and Olivier Ratheaux. The results of our research related to the Effects Method are summarised in Bombart; Palenberg (2010): Friend or Foe? The Effects Method and Cost-Benefit Analysis. Munich: Institute for Development Strategy.

First Name	Last Name	Job description	Home institution	Interview modality
Niels	Keijzer	Programme Officer	European Centre for Development Policy Management	Telephone
Jeannette	Monier	Evaluation Officer	European Commission	Face to face
Jan	Soeltenfuss	Counsellor Trade & Investment	European Investment Bank	Telephone
Tullia	Aiazzi	Evaluation Officer	Food and Agriculture Organization	Face to face
Marianne	Reuber	Senior Policy Advisor, Evaluation Department	Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ)	Face to face
Martina	Vahlhaus	Director, Evaluation Department	Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ)	Face to face
Luigi	Cuna	Evaluation Officer	International Fund for Agricultural Development (IFAD)	Face to face
Michael	Flint	Consultant	International Fund for Agricultural Development (IFAD)	Face to face
Shyam	Khadka	Senior Portfolio Manager	International Fund for Agricultural Development (IFAD)	Face to face
Luciano	Lavizzari	Director, Office of Evaluation	International Fund for Agricultural Development (IFAD)	Face to face
Ashwani	Muthoo	Senior Evaluation Officer	International Fund for Agricultural Development (IFAD)	Face to face
Michael	Bamberger	Consultant	Independent	Telephone
Pedro	Belli	Consultant	Independent	Telephone
Michel	Le Gall	Consultant	Independent	Face to face
John	Mayne	Consultant	Independent	Telephone
Michael	Patton	Consultant	Independent	Exchange of emails
Sandra	Speer	Consultant	Independent	Telephone

First Name	Last Name	Job description	Home institution	Interview modality
Philippe	Hugon	Professor of Economy and Director Research	Institut de Recherche Internationales et Stratégiques (IRIS)	Face to face
Sonja	Nelles	Senior Project Management Quality and Evaluation	InWEnt (Capacity Building International)	Face to face
Theodor	Dickmann	Evaluation Officer (at time of interview)	KfW Entwicklungsbank	Face to face
Marlis	Sieburger	Evaluation Officer	KfW Entwicklungsbank	Face to face
Eva	Terberger	Director, independent Evaluation Department	KfW Entwicklungsbank	Face to face
Franck	Wiebe	Chief Economist	Millennium Challenge Corporation	Face to face
Stefan	Calvi	Partner	PricewaterhouseCoopers AG Wirtschaftsprüfungsgesellschaft	Telephone
Aliza	Belman Inbal	Director, International Development Program	Tel Aviv University's School of Government and Policy	Telephone
Johannes	Dobinger	Evaluation Officer	United Nations Industrial Development Organization (UNIDO)	Telephone
Robert	Schwartz	Director of Evaluation and Monitoring	The Ontario Tobacco Research Unit, University of Toronto	Telephone
Xavier	Legrain	Director (Acting) Quality Assurance Group (QAG)	World Bank	Face to face
Gary	McMahon	Consultant	World Bank	Email exchange
Soniya	Carvalho	Lead Evaluation Officer	World Bank Independent Evaluation Group	Face to face
Christopher	Gerrard	Lead Evaluation Officer	World Bank Independent Evaluation Group	Face to face
Andrew	Warner	Lead Economist	World Bank Independent Evaluation Group	Face to face

First Name	Last Name	Job description	Home institution	Interview modality
Han	Fraeters	Manager, Knowledge Exchange Practice	World Bank Institute	Face to face
Samuel	Otoo	Manager, Capacity Development and Results Practice	World Bank Institute	Face to face

Annex B: Bibliography

Adler, M. D.; Posner, E. A. (2006): New Foundations of Cost-Benefit Analysis. Cambridge: Harvard University Press

Arrow, K. J. (1951): Social Choice and Individual Values. New York: Wiley

Averch, H. A. (2004): Using Expert Judgment. Handbook of Practical Program Evaluation. J. S. Wholey, H. P. Hatry and K. E. Newcomer. San Francisco: Jossey-Bass

Balassa, B. (1976): The "Effects Method" of Project Evaluation once again. Oxford Bulletin of Economics and Statistics 39

Balassa, B. (1976): The "Effects Method" of Project Evaluation. Oxford Bulletin of Economics and Statistics 38

Bamberger, M.; Rugh, J.; Mabry, L. (2006): RealWorld Evaluation: Working Under Budget, Time, Data, and Political Constraints. Thousand Oaks: Sage Publications

Belli, P.; Anderson, J.; Barnum, H.; Dixon, J.; Tan, J.-P. (1997): Handbook on Economic Analysis of Investment Operations. Washington, D.C.: The World Bank

Bentlage, A.; Petry, N. (2008): Einzelwirtschaftliche Investitionsanalyse. Frankfurt am Main: KFW Entwicklungsbank

BMF (2001): Allgemeine Verwaltungsvorschriften (VV-BHO), VV-BHO §7, H 05 01. Bundeshaushaltsordnung (BHO): BMF

BMZ (2001): Hauptbericht über die Evaluierung: Berücksichtigung von Wirtschaftlichkeitsüberlegungen bei EZ-Vorhaben. Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung

Boardman, A. E. (2006): Cost-Benefit Analysis : Concepts and Practice. Upper Saddle River, N.J.: Pearson/Prentice Hall

Bombart, D.; Palenberg, M. (2010): Friend or Foe? The Effects Method and Cost-Benefit Analysis. Munich: Institute for Development Strategy

Brent, R. J. (2006): Applied Cost-Benefit Analysis. Cheltenham, UK ; Northampton, MA: Edward Elgar

Campbell, H. F.; Brown, R. P. C. (2003): Benefit-Cost Analysis: Financial and Economic Appraisal Using Spreadsheets. Cambridge, UK ; New York: Cambridge University Press

Castillo, C. d. (1998): Economic Analysis of Social Investment Fund Projects. Washington, D.C.: World Bank

Chan, S. (1982): Expert Judgments under Uncertainty: Some Evidence and Suggestions. Social Science Quarterly 63

Chervel, M. (1977): The Rationale of the Effects Method: a Reply to Bela Balassa. Oxford Bulletin of Economics and Statistics 39

Chervel, M. (2008): Mondialisation des Marchandises et Marchandisation du Monde: Matériau de Pensée Economique pour le Nouveau Siècle. Paris

Chowdhury, A.; Kirkpatrick, C. H. (1994): Development Policy and Planning : an Introduction to Models and Techniques. London ; New York: Routledge

Copeland, T. E.; Koller, T.; Murrin, J. (2000): Valuation: Measuring and Managing the Value of Companies. Wiley Frontiers in Finance. New York: John Wiley

Davies, R.; Dart, J. (2005): The 'Most Significant Change' (MSC) Technique: A Guide to Its Use Version 1.0

Dinwiddy, C. L.; Teal, F. (1996): Principles of Cost-Benefit Analysis for Developing Countries. Cambridge; New York: Cambridge University Press

Dreze, J.; Stern, N. (1994): Shadow Prices and Markets: Policy Reform. Cost-Benefit Analysis. P. R. G. Layard and S. Glaister

Duflo, E.; Kremer, M. (2003): Use of Randomization in the Evaluation of Development Effectiveness. World Bank Operations Evaluation Department Conference on Evaluation and Development Effectiveness. Washington, D.C.

Eisner, E. W. (1985): The Art of Educational Evaluation : a Personal View. London; Philadelphia: Falmer Press

Eisner, E. W. (2004): The Roots of Connoisseurship and Criticism - A Personal Journey. Evaluation Roots - Tracing Theorists' Views and Influences. M. C. Alkin. Thousands Oaks, CA: Sage Publications

Eureval-3CE; DG Budget (2006): Study on the Use of Cost-Effectiveness Analysis in EC's Evaluations. Lyon: Centre for European Expertise and Evaluation

European Commission (1997): ECO-Fin Manual. Brussels: European Commission

European Commission (2008): Guide to Cost-Benefit Analysis of Investment Projects. Brussels: European Commission

European Commission (2009): Results Oriented Monitoring: ROM Handbook. Brussels: European Commission

Forss, K.; Vedung, E.; Kruse, S. E.; Mwaiselage, A.; Nilsdotter, A. (2008): Are Sida Evaluations Good Enough? An Assessment of 34 Evaluation Reports. Sida Studies in Evaluation. Stockholm: Swedish International Development Cooperation Agency

Franck, B. (1996): The Effects Method and Economic Cost-Benefit Analysis: Substitutes or Complements? Cost-Benefit Analysis and Project Appraisal in Developing Countries. C. H. Kirkpatrick and J. Weiss. Cheltenham, UK ; Brookfield, Vt., US: E. Elgar

Franklin, B.; Labaree, L. W.; Bell, W. J.; Poor Richard Club (Philadelphia Pa.); Rogers, B.; Benjamin Franklin Collection (Library of Congress); Pforzheimer Bruce Rogers Collection (Library of Congress) (1956): Mr. Franklin : a Selection from his Personal Letters. New Haven, London: Yale University Press; Geoffrey Cumberlege, Oxford University Press

Frechtling, J. A. (2007): Logic Modeling Methods in Program Evaluation. San Francisco: Jossey-Bass

Fuguitt, D.; Wilcox, S. J. (1999): Cost-Benefit Analysis for Public Sector Decision Makers. Westport, Conn.: Quorum

Gramlich, E. M. (1990): A Guide to Benefit-Cost Analysis. Englewood Cliffs, N.J.: Prentice-Hall

Greenberg, D. H.; Appenzeller, U. (1998): Cost Analysis Step by Step : a How-to Guide for Planners and Providers of Welfare-to-Work and other Employment and Training Programs. New York, N.Y.: Manpower Demonstration Research Corporation

Greiner, J. M. (2004): Trained Observer Ratings. Handbook of Practical Program Evaluation. J. S. Wholey, H. P. Hatry and K. E. Newcomer. San Francisco: Jossey-Bass

GTZ (1993): Anleitung für die ökonomische Beurteilung von Projekten der Technischen Entwicklungszusammenarbeit. Frankfurt am Main: GTZ

Guba, E. G.; Lincoln, Y. S. (1989): Fourth Generation Evaluation. Newbury Park, Calif.: Sage Publications

Hammond, M. A. (2006): A Framework for Evaluating Budget Support. S. Koeberle, Z. Stavreski and J. Walliser. Washington, D.C.: World Bank

Hearn, S.; Jones, H. (2009): Outcome Mapping: a Realistic Alternative for Planning, Monitoring and Evaluation. London: Overseas Development Institute

Hedblom, E.; Hildenwall, V. (2008): Strengthening the Results-Orientation in Sweden's Engagement in Multilateral Development Cooperation: An Evaluation of UNDP's Country Level Evaluation Activities. Karlstad: Swedish Agency for Development Evaluation

Hemmer, H.-R. (2005): Einführung in die Cost-Benefit Analyse. Frankfurt am Main: KfW Entwicklungsbank

Hildred, W.; Beauvais, F. (1995): An Instrumentalist Critique of Cost-Utility Analysis. Journal of Economic Issues

Hirschman, A. O. (1995): Development Projects Observed. Washington, D.C.: Brookings Institution

Hoffmeister, W. (2008): Investitionsrechnung und Nutzwertanalyse. Berlin: Berliner Wissenschafts-Verlag

IFAD (2009): Evaluation Manual: Methodology and Processes. Rome: International Fund for Agricutural Development, Office of Evaluation

Kee, J. E. (2004): Cost-Effectiveness and Cost-Benefit Analysis. Handbook of Practical Program Evaluation. J. S. Wholey, H. P. Hatry and K. E. Newcomer. San Francisco: Jossey-Bass

Keeney, R. L.; Raiffa, H. (1993): Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge England; New York, NY, USA: Cambridge University Press

Kirkpatrick, C. H.; Weiss, J. (1996): Cost-Benefit Analysis and Project Appraisal in Developing Countries. Cheltenham, UK; Brookfield, Vt., US: E. Elgar

Langhammer, A.; Nägele, E. (2005): Einführung in die Gesamtwirtschaftliche Investitionsanalyse. Frankfurt am Main: KfW Entwicklungsbank

Levin, H. M.; McEwan, P. J. (2001): Cost-Effectiveness Analysis: Methods and Applications. Thousand Oaks, Calif.: Sage Publications

Little, I. M. D.; Mirrlees, J. A. (1974): Project Appraisal and Planning for Developing Countries. New York: Basic Books

Little, I. M. D.; Mirrlees, J. A. (1994): The Costs and Benefits of Analysis: Project Appraisal and Planning Twenty Years on. Cost-Benefit Analysis. P. R. G. Layard and S. Glaister. Cambridge: Cambridge University Press

Mathers, C.; Fat, D. M.; Boerma, J. T. (2008): The Global Burden of Disease: 2004 Update. Geneva, Switzerland: World Health Organization

Mayne, J. (2001): Adressing Attribution through Contribution Analysis: Using Performance Measures Sensibly. Canadian Journal of Program Evaluation

McMahon, G. (1997): Applying Economic Analysis to Technical Assistance Projects. Policy Research Working Paper Washingotn, D.C.: World Bank 1749

Meimberg, R. (1971): Voraussetzungen einer globalen Entwicklungspolitik und Beiträge zur Kosten- und Nutzenanalyse. Berlin: Dunker&Humblot

Ministère de la Coopération (1978): The Methodology of Planning, Manual of Economic Evaluation of Projects, The Effects Method. Paris: Ministère de la Coopération

Mishan, E. J.; Quah, E. (2007): Cost-Benefit Analysis. London; New York: Routledge

Nas, T. F. (1996): Cost-Benefit Analysis: Theory and Application. Thousand Oaks, Calif.: Sage Publications

OECD DAC (2002): OECD DAC Glossary of Key Terms in Evaluation and Results-Based Management. Paris: OECD Development Assistance Committee

OECD DAC (2010): Evaluating Development Co-Operation: Summary of Key Norms and Standards. Paris: OECD

Palenberg, M. (2009): Tools and Methods for Evaluating Aid Efficiency: Inception Report. Bonn: BMZ

Patton, M. Q. (2002): Qualitative Research and Evaluation Methods. Thousand Oaks, Calif.: Sage Publications

Patton, M. Q. (2008): Utilization-Focused Evaluation. Thousand Oaks: Sage Publications

Pearl, J. (2009): Causality: Models, Reasoning, and Inference. Cambridge; New York: Cambridge University Press

Piveteau, A. (2004): Évaluer les ONG. Paris: Edition Karthala

Price, C. (1996): Discounting and Project Appraisal: From the Bizarre to the Ridiculous. Cost-Benefit Analysis and Project Appraisal in Developing Countries. C. H. Kirkpatrick and J. Weiss. Cheltenham, UK; Brookfielt, Vt., US: E.Elgar

Rogers, P. J. (2000): Program Theory in Evaluation: Challenges and Opportunities. San Francisco, Calif.: Jossey-Bass

Rossi, P. H.; Lipsey, M. W.; Freeman, H. E. (2004): Evaluation : a Systematic Approach. Thousand Oaks, CA: Sage

Shadish, W. R.; Cook, T. D.; Campbell, D. T. (2002): Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin

Squire, L.; Tak, H. G. v. d. (1975): Economic Analysis of Projects. A World Bank Research Publication. Baltimore: Published for the World Bank by Johns Hopkins University Press

Stokey, E.; Zeckhauser, R. (1978): A Primer for Policy Analysis. New York: W. W. Norton

Stufflebeam, D. L.; Shinkfield, A. J. (2007): Evaluation Theory, Models, and Applications. San Francisco: Jossey-Bass

The White House (1993): Executive Order 12866 of September 30. Federal Register, Regulatory Planning and Review 58

UNIDO (1972): Guidelines for Project Evaluation. New York: UNIDO

Valadez, J. J.; Bamberger, M. (1994): Monitoring and Evaluating Social Programs in Developing Countries: a Handbook for Policymakers, Managers, and Researchers. EDI Development Studies. Washington, D.C.: World Bank

Voß, H. (1993): Die Wirtschaftlichkeitskontrolle von Entwicklungsprojekten. Frankfurt am Main: Johann Wolfgang Goethe Universität Diplom

Weirich, P. (2001): Decision Space: Multidimensional Utility Analysis. Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge, UK; New York: Cambridge University Press

Weiss, C. H. (1998): Evaluation: Methods for Studying Programs and Policies. Upper Saddle River, N.J.: Prentice Hall

World Bank (1994): The World Bank Operational Manual, Operational Policies, Economic Evaluation of Investment Operations. Washington: World Bank

World Bank Independent Evaluation Group (2009): Annual Review of Development Effectiveness 2009: Achieving Sustainable Development. Washington: World Bank

World Bank Independent Evaluation Group (2010): Cost-Benefit Analysis in World Bank Projects. Washington: World Bank

Yoon, K.; Hwang, C. L. (1995): Multiple Attribute Decision Making : an Introduction. Sage University Papers Series. Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage Publications 07

Zangenmeister, C. (1970): Nutzwertanalyse in der Systemtechnik. München: Verlagskommission Wittmannsche Buchhandlung