Thematic Network

‚Policy and Project Evaluation Methodologies‘

TRANS-TALK

Final Report

Table of Contents

1 EXECUTIVE SUMMARY 1

2 OBJECTIVES 6

3 WORK DESCRIPTION 8

4 RECOMMENDATIONS 23

5 ANNEX A. SUMMARY OF WORKSHOP PROCEEDINGS 32

6 ANNEX B. GUIDELINES FOR IMPROVING EVALUATION PRACTICES IN TRANSPORT 68

7 ANNEX C. TRANS-TALK DOCUMENTATION 92

8 REFERENCES 93

9 CONTACT DETAILS 100

Thematic Network ‚Policy and Project Evaluation Methodologies‘

TRANS-TALK

Final Report

1 Executive Summary

The TRANS-TALK Thematic Network was set up in January 2000 under the Fifth Framework Programme with the objective to provide a networking platform for those involved in the field of transport evaluation; explore the conceptual and empirical problems in contemporary transport evaluation; and develop guidelines that help improve transport evaluation.

The launch of the TRANS-TALK thematic network reflects the renewed strategic interest in transport evaluation. This has different sources – conceptual, empirical and policy-relevant.

At the conceptual level, this renewed interest in transport evaluation is related to the realisation that the conventional methods for evaluation that apply to infrastructure appraisal, like cost-benefit analysis, are not adequate to address contemporary challenges to transport policy. At the empirical level, there is, on the one hand, the problem of choice among a multitude of methodologies and, on the other, the problem of measurement and comparison. Finally at the policy-relevant level, there is today an intensified demand for evaluation that looks at the (long-term) outcomes of policy rather than alone its (short-term) direct outputs.

Current state and main trends. Transport-specific evaluation frameworks exist in most Member States of the European Union, albeit mainly for infrastructure assessment at project level. These are used for ex-ante evaluation or appraisal, and for the prioritising and phasing of projects.

The appraisal of transport infrastructure relies mostly on cost-benefit (CBA) and multi-criteria analysis (MCA). Road and rail projects are the ones most commonly subjected to CBA. Analytically CBA is widely used for the so-called direct transport impacts. Among environmental impacts, noise and local air pollution are included in appraisal across Member States but they are valued in monetary terms only in half. The treatment of indirect socio-economic impacts is uneven.

At the programme level, and with few exceptions, the drawing of transport master plans is common practice in EU Member States at least for one mode –

most frequently again road. Policy or multi-modal master plans, on the other hand, are less common, in part because of the remaining co-ordination problems between relevant agencies or policy institutions.

At the European level we find no integrated or harmonised evaluation framework to apply to projects, programmes or policies that are of common interest or added value. This is in part the result of the lack of harmonisation with regard to transport data, forecasts, models or scenarios. The other reason for the absence of a harmonised evaluation framework is largely political, and related to the demand for flexibility by Member States in view of the subsidiarity principle.

Recommendations:

At the policy level there are two types of evaluation that are relevant, namely, output evaluation at regular intervals to examine or check whether the action points or measures announced have also been implemented; and outcome evaluation at the short-, medium- and long-term for the purpose of monitoring with reference to broad impact indicators. A combined monitoring system combining output- and outcome-specific indicators is a support tool that no decision-making level can do without, if policy evaluation is to be taken seriously.

The level of detail of the evaluation of a programme's results or outcomes depends on the character of the programme. Some programmes are so multifarious that evaluation can only be carried out at the aggregate level, as with policy. Other programmes are more clearly delineated, thus allowing the application of more detailed evaluation techniques.

At the project level, a certain degree of harmonisation is called for with regard to the background assumptions to guide forecasting exercises. The same applies to transport data and documentation.

In view of the conflicting interests whether there should be (or not) a harmonised evaluation framework for transport, the open co-ordination method should here be used as prescribed by the recent White Paper on European Governance.

Pressures for change. The gradual consolidation of the European Union as a polity, with own institutions, an own redistribution budget, own policy networks and agendas and own decision processes, represents a challenge to evaluation in the European context. This is not least because of the subsidiarity principle which prescribes that the Union may only intervene where there would be an added value from this intervention. Only there does not always exist either clarity or agreement as to how to define and measure this added value.

At the more technical level of information, being able to evaluate the European-added value also means being able to rely on at least a minimum set of harmonised data inputs, and, in part, modelling outputs, including projections and forecasts. Data availability and comparable information remains a problem. Furthermore, our knowledge on the underlying travel behaviour of individuals and of companies and how this might differ across countries is still not adequate.

At a different level, the process of European integration has tended to strengthen the deliberative aspects of evaluation. In the European context and in view of the character itself of the Union, evaluation is more often than not seen as part of a negotiation and deliberation process, through which socially desirable transport actions are gradually identified. This implies promoting both stakeholder involvement in evaluation and a stronger emphasis on transparency and accountability.

Recommendations:

The evaluator must ask and document whether a policy, programme or project was implemented and/or how successful it was, yet must also inquire into those factors, specific to the decision context, that have facilitated or obstructed its implementation.

Complex decision-making procedures call for a new culture of evaluation that places more emphasis on deliberation, especially with regard to appraisal. At the minimum level this involves making conceptual and methodological choices transparent to allow reflective discussion – making evaluations open for critical review is a major step towards establishing a deliberation culture in evaluation. At the maximum level, the evaluation should be thought of as simulating a decision process and use methods for gathering different opinions and bringing different actors together to discuss issues of common concern.

A multiplicity of methods. There are several methods or tools that can be used for evaluation. These can be classified according to four dimensions: (a) the analytical framework from which they emerged; (b) their suitability for policy, programme or project evaluation; (c) their suitability for different phases of policy analysis, and (d) the extent to which they rely on statistical or mathematical methods or models, or alternatively make use of quantitative methods.

Several problems can arise by failing to be explicit about the underlying assumptions of analytical frameworks used in evaluation. One frequent result is the subsequent failure to be clear about the objectives of evaluation exercises. Another possible outcome are serious misunderstandings between those who commission and those who are assigned the task of evaluation. Perhaps the most serious problem is the failure to understand or rightly interpret the results

of an evaluation exercise and, in the case of multi-tier evaluation exercise, to integrate the various results.

Recommendations:

The multiplicity of evaluation methods can be mistaken to imply that everything goes, more specifically that either any (desirable) result is possible to deliver if the choice of methods is carefully thought through, or that differences in evaluation results are methodologically grounded. Absolute relativism is however not the conclusion to draw from the multiplicity of methodologies. Robust findings will remain even when using different methodologies.

Quality standards apply with regard to the use of methods. Failure to apply these standards can lead to distorted evaluation results.

In dealing with uncertainty of estimations in evaluation, sensitivity analysis and the use of scenarios is recommendable.

Technical challenges. Scientific inquiry has progressed with regard to several contemporary difficult issues for transport evaluation, like transport and economic development; network effects and European-added value; time, term and uncertainty; the welfare basis of evaluation; and environmental valuation.

Some open questions remain. These are in part technical, in part conceptual. Disagreement or lack of clarity with regard to concepts and definitions of terms have, therefore, to be resolved in advance.

Recommendations:

The review of contemporary evaluation challenges in transport has revealed a series of subjects for which research is needed.

For some areas, the type of research which is currently needed is more of a technical, development or demonstration type. In other areas, there is a lack of either empirical information or of conceptual elaboration. Research programmes need to distinguish between these two types of research priorities.

A better integration of technical and political perspectives. Transport evaluation has till now been mainly technocratic in focus. As a result, there is a weak link between the community of transport professionals who undertake the analysis and decision-makers or their policy advisers, which can result in communication failure.

The role of the evaluator is in that problematic because it represents (or ought to represent) an intermediary position between the politician who decides and presents, and the scientist who studies and criticises. Therefore important for

evaluation perceived as a process is that different functional roles are represented through the setting up of appropriate teams.

The political interest in evaluation is only a problem if it is one-sided, reflecting vested interests or not accountable. This is also why evaluation ought not to be the monopoly of government or dependent on governmental or state funding. It should also not be primarily or solely carried out internally.

Finally there are two other important factors for promoting the better integration of political and technical perspectives in evaluation. The first is realism – evaluation is a science of interpretation and in that neither a substitute for decision nor for deliberation. The second factor is organisational and reflects the fact that evaluation is a process. Just as it is important for evaluation teams to reflect on the production and dissemination of knowledge, recognising that these are two different functions that might also have to be separated at the personal level, commissioning agencies need to recognise that evaluation does not end with the commissioning of a study but that it is important to install procedures that ensure that the evaluation results are monitored and, once available, reach those that need them in the most appropriate format.

Recommendations:

The better integration of technical and political perspectives in evaluation requires changes from both the technical and from the political side. More realism and a greater willingness to co-operate are minimum requirements. Beyond this, the different functions of evaluation, on the one hand, and of politics on the other, have to be recognised and respected.

2 Objectives

The TRANS-TALK Thematic Network was set up in January 2000 under the Fifth Framework Programme (Key Action 'Sustainable Mobility and Intermodality) with the following objectives:

- To provide a networking platform for those involved in the field of transport evaluation from either the demand or supply sides;
- To explore the conceptual and empirical problems in contemporary transport evaluation;
- To develop guidelines that help improve transport evaluation in providing a preliminary framework for the integration of different policy and project evaluation methodologies, thus setting standards in the field.

The launch of the TRANS-TALK thematic network reflects the renewed strategic interest in transport evaluation. This has different sources:

At the *conceptual level*, it results from the realisation that the conventional methods for evaluation and transport planning that apply to infrastructure assessment, like cost-benefit analysis, are not adequate in their present form to address issues that relate to transport policy, but also not sufficient to measure the achievement of objectives like environmental sustainability or social cohesion that are also relevant for infrastructure investment.

At the *empirical level*, there are two sets of problems that explain this renewed interest in the principles and application of transport evaluation. On the one hand, there is the problem of choice among a multitude of methodologies. Even when conceptually similar, these can be different with regard to their underlying assumptions, the valuation techniques or procedures used or the range of impacts covered. This, for instance, is the main problem surrounding the choice of appropriate traffic or transport models for making forecasts or simulating policy impacts. On the other hand, the analyst is confronted with several measurement problems: good quality *and* comparable transport data, including on standard issues like safety, road links or traffic counts, are the exception rather than the rule.

Finally at the *policy level*, there is today an intensified demand for evaluation that looks at the (long-term) outcomes of policy rather than alone its (short-term) direct outputs. Also in transport there are increased demands for 'strategic assessments' for checking the consistency of 'policies, plans and programmes (P/P/P)' or for approaching 'the design of projects in a generic sense' (EC, 1994; EC, 1997a). A recent European Initiative – the *Sound and Efficient Management 2000 Programme* – has sought to elaborate a framework for carrying out evaluations of Community programmes (cf. EC, 1997b; EC, 1999)

and has increased awareness of the significance of the evaluation function in policy formulation and deliberation.

3 Work Description

3.1 Three workshops and supporting activities

As a thematic network, the work programme of TRANS-TALK was implemented through three workshops and supporting activities that included literature reviews as input to the setting of the workshop agendas, the publication of the conference proceedings, the writing up of synthesis reports to summarise the conclusions and main recommendations deriving from each workshop, and a survey among transport RTD projects to obtain feedback on the themes of the thematic network and for setting up a Virtual Library (www.iccr-international.org/trans-talk/library)

The three workshops organised in the framework of TRANS-TALK were:

- WS1: Policy and Project Evaluation: Context, Theory and Methods
  May 29th to May 31st, 2000 Brussels

- WS2: Projects, Programmes, Policies: Evaluation Needs and Capabilities
  November 6th to November 8th, 2000 Brussels

- WS3: Improving Evaluation Practices in Transport: Towards a Better Integration of Technical and Political Perspectives
  May 30th to June 1st, 2001 Brussels

The papers presented at the workshops are available for downloading at the network's Web Site at www.iccr-international.org/trans-talk. Revised updated versions of these papers can be read in the two TRANS-TALK book publications forthcoming with the Avebury Publishers of the Ashgate Publishing House (Giorgi *et al.*, eds., 2000; Pearman *et al.*, eds., 2001).

Summaries of the presentations and discussions were prepared for the regular progress reports and project's deliverables and are reproduced in Annex A of this final report.

The guidelines for improving evaluation practices in transport prepared as input to the third workshop were submitted to a consultation procedure using the Internet (www.iccr-international.org/trans-talk/consultation) and discussed at length at the third workshop. The revised version of these guidelines can be read in Annex B of this final report.

Annex C of this report lists the TRANS-TALK documentation and deliverables.

Below we discuss the main findings of TRANS-TALK by considering:

- first, the current state of transport evaluation;
- second, the pressures for change, including those specific to the decision-making context;
- third, the multitude of evaluation methodologies and the problems this poses;
- fourth, the main technical challenges to transport evaluation, and
- fifth, the integration of technical and political perspectives in evaluation.

Our conclusions and recommendations are based on the review of and reflection on the various contributions to the three TRANS-TALK workshops, a summary of which, as already indicated, can be read in Annex A.

3.2 The current state and main trends in transport evaluation

Generic frameworks of evaluation include those developed by international organisations like the World Bank or the OECD, those elaborated by the Commission in the framework of the Sound and Efficient Management Initiative, SEM 2000, and those developed by national administrations, like the U.S. and the U.K. These generic frameworks come usually in the form of guidelines and consider evaluation principles and design. Deliverable 2 of TRANS-TALK *The Theory and Practice of Evaluation; Conclusions from the First Workshop* reviewed and summarised several of these frameworks. A representative list of bibliographical references was provided with the TRANS-TALK *Guidelines for Improving Evaluation Practices in Transport* and is reproduced in the bibliography of this report.

Transport-specific evaluation frameworks exist in most Member States of the European Union, albeit mainly for infrastructure assessment at **project level**. These are used for ex-ante evaluation or appraisal, and for the prioritising and phasing of projects.

The EUNET project (1996 – 2000) funded under the Transport RTD Programme of the Fourth Framework carried out a review of current appraisal practice in EU Member States. The results of this review are reported in Deliverable 9 of EUNET, *Measurement and Valuation of the Impacts of Transport Initiatives*, Version 2.1, 1998 (authors: John Nellthorp, Peter Mackie and Abigail Bristow) and were also presented and discussed at the 1st TRANS-TALK workshop in November 2000 with several Member State representatives. A detailed analysis of the results of this review, including a comparison of the technical elements of several appraisal methods, can also be read in S. M. Grant-Muller, P. Mackie, J. Nellthorp and A. Pearman (2001), 'Economic Appraisal of European Transport

Projects: The State-of-the-Art Revisited', *Transport Reviews*, Vol. 21, No. 2, pp.237-261.

This review showed that the appraisal of transport infrastructure relies mostly on cost-benefit (CBA) and multi-criteria analysis (MCA). All EU Member States (with the exception of Luxembourg) use CBA on at least one mode in the transport infrastructure planning process. Road and rail projects are the ones most commonly subjected to CBA. This shows the importance attached to CBA as a value-for-money measure in appraisal. However, the incomplete nature of most CBAs, especially with regard to the valuing of environmental or social costs, means that in practice the CBA result is taken as one among several inputs to the decision and is thus not dominant. This is especially the case for air, sea and inland waterways where CBAs are also much less common. More generally, in most countries the overall appraisal embraces not only the CBA result but also some form of qualitative appraisal of the social, economic and environmental effects.

Analytically, CBA is widely used for the so-called direct transport impacts, in particular construction costs, vehicle operating costs, times savings and safety. Among environmental impacts noise and local air pollution are included in appraisal across Members States but they are valued and included in CBA only in around half of these. Finally, the treatment of indirect socio-economic impacts is uneven. Nevertheless there is widespread recognition that conventional CBA omits this category of impacts, yet that it is important to consider these through other means, mainly in a multi-criteria assessment framework.

At the **programme level**, and with few exceptions, the drawing of transport master plans is a common practice in EU Member States at least for one mode – most frequently road. Policy or multi-modal masterplans, on the other hand, are less common.

Characteristic of the master plans recently published to cover the first decade of the 21st century is their convergence in terms of objectives – a fact not unrelated to the intensification of the European project of integration – and the explicit intention to introduce stricter monitoring mechanisms for evaluating their implementation. The latter is also a fixed element of the new *White Paper on the Common Transport Policy*.

Deliverable 5 of TRANS-TALK *Projects, Programmes, Policies: Evaluation Needs and Capabilities* includes a review of masterplan practice in select EU member States – more specifically the Scandinavian countries, the Netherlands, France, UK, Italy and Germany. This review has revealed the following cross-cutting themes.

1. The planning process relies today more on deliberation than it did earlier. Especially with regard to multi-modal planning, co-ordination among agencies and/or departments across Ministries would seem unavoidable.

2. Master plans rely on projections and forecasts. The underlying model represents a formulation of major links between the transport system and its socio-economic environment or between transport system components. It can be a conceptual or statistical model. In most countries both approaches are used with a varying degree of emphasis on econometric models.

3. Often there is no single model used to produce projects and/or forecasts. Instead several models are combined – not necessarily consistently across time (or projects) – to delineate an appropriate quantitative framework. The problem here is that it is difficult, if not impossible, to identify which model has been used to produce what output and what simply derives from a qualitative expert analysis. More transparent and systematic documentation procedures are, therefore, called for.[1]

4. The lack of appropriate and harmonised transport statistics between modes (and at European level also between countries) is constantly cited as hampering the proper development of the planning process. The establishment of the European Transport Information System (ETIS) in the next few years is expected to greatly facilitate matters in this respect.

5. Scenario techniques are increasingly used in modal and multi-modal planning. They have also become more sophisticated, no longer relying on GDP growth projections alone. In the Netherlands scenarios include hypotheses about structural and/or behavioural changes; in France and Germany they consider political choices, whereas in Norway and Sweden they incorporate distinct transport policy strategies. The sharing of scenario information between countries remains however the exception rather than the rule. This makes strategic assessment at European level all the more difficult.

6. With regard to traffic, few countries make network assignments and when they do, none seem to give a very precise overview of origin-destination exchanges between European regions. With the exception of the Netherlands, the network assignment of international traffic is rarely assessed in detail. At a time when international traffic grows faster than national traffic, this is a significant omission. The modelling of operating

[1] This problem is, among others, examined by the THINK-UP thematic network (www.netr.fr/think-up). With regard to European models, the SPOTLIGHTS thematic network is currently in the process of developing a platform for the exchange of transportation modelling information (http://gtf.mkm.de).

systems as well as of interactions between transport and land use likewise receives little attention.

Programmatic evaluation in the transport sector is generally becoming more commonplace due to the drive towards a better cross-sectoral integration of policies. This is reflected in the various national master plans which are more outcome oriented than before, but also in the discussions surrounding the introduction of strategic environmental assessment (SEA) and the renew of interest in ex-post evaluations.

At the **European level** we find no integrated or harmonised evaluation framework to apply to projects, programmes or policies that are of common interest or added value. Beginning with the EURET programme in the early nineties, several RTD projects developed assessment methods or models to apply at European level. None of these however could develop into a standard methodology.

Should transport evaluation be standardised and harmonised at European level? For instance, should all Member States be called to use not only the same methods in project appraisal but also the same indicators / thresholds etc.? This remains an open question. There are arguments in favour and arguments against such an approach.

What speaks in favour of harmonisation at European level are the existing difficulties with comparability (of forecasts, projections, assessments of multilateral activities, etc.) and, resulting from this, the frequent failure of assessments carried out at national level to sufficiently reflect the European added value or network effects. This is the position advocated primarily by the Commission.

What speaks against harmonisation is the legitimate variation at the national level (but also more generally) of policy objectives, on the one hand, and of scientific conventions, on the other. Hence also the demands for flexibility and the worries that extensive harmonisation might not be consistent with democracy. This is the position advocated by the majority of Member States.

The debate on harmonisation vs. flexibility is not specific to transport evaluation. Similar discussions are in fact currently taking place across policy sectors and, not least importantly, more generally with regard to the European polity and its future.

In the case of transport evaluation, the two approaches need not be seen as exclusive. Allowing for flexibility (for Member States) does not exclude that the Commission proceeds to establish its own standard procedures / methods. More importantly perhaps, it does not exclude seeking harmonisation at a basic minimum level.

An integrated generic framework for transport evaluation is desirable to the extent that it helps clarify perspectives and approaches. It is possible in the form of general guidelines (like the TRANS-TALK guidelines) that establish procedures and a minimum set of standards or conventions with regard to methods or data.

More detailed standards or conventions must be tailored to the needs of specific sub-sectors or areas. The needs or concerns of safety related transport evaluation are, for instance, different from those of infrastructure project evaluation. Here too, however, it is possible and desirable to seek through deliberation to establish ground rules. Such ground rules can serve integration by facilitating co-ordination and comparison, yet need not imply full harmonisation and the loss of flexibility.

3.3 Pressures for change: the decision-making context

The gradual consolidation of the European Union as a polity with own institutions, an own redistribution budget, own policy networks and agendas and own democratic decision processes – in brief the superseding of the nation-state government structures through a supra-national entity – represents a challenge to evaluation carried out for European institutions or in the European context.

The reason for this has to do with the nature of the European polity. The European Union has assumed many functions of a (nation-) state yet its sovereignty is constructed according to the principle of subsidiarity. This requires the continuous negotiation and clarification of the legitimate areas of policy intervention for the European Union (through legislation, economic or other measures).

The term European-added value is used to delineate such areas of intervention. Policy interventions from the European level are justified if there can be expected an added value from such interventions. Accordingly, the degree or type of intervention, but also the question of responsibility and accountability for a decision, and the monitoring of its implementation – not least the financial implications, i.e. who pays and how much – can be determined. Evaluations carried out for European institutions are often called to relate to this added value or indeed operationalise it in their design and assessments.

That this is a far more problematic area than might appear at first sight is shown by the example of the TEN-T. Trans-national infrastructure investment plans aiming to achieve an efficient transport network for the whole of Europe are often operationalised in terms of ‘key links’ which in turn represent specific infrastructure projects in particular countries (for instance the Brenner tunnel) or

cross-boundary projects (like the Øresund link). The evaluation of such projects is typically carried out at national level, often using national parameters that might not be adequate for measuring the degree of strategic impacts on a wider scale, or what is called network effects. Or if they do, the implications of the analysis with regard to the fair distribution of costs (according to benefits) are not considered. The failure of several TEN-T projects to be implemented reflects the failure to consider operationally their European-added value in the policy process and subsequently in their evaluation.

This problem is perhaps not very different from that faced in strongly federalised countries where regions have a high degree of autonomy with regard to the decision and financing of projects through their area of jurisdiction. However unlike at the European level, the national-regional governance interface is characterised by a sharing of a minimum set of common parameters for evaluation, and flexible rules about sharing financing burdens. Furthermore, in relative terms the re-distributive budget available to nation-states is far higher than that available to the European Union.

It would therefore seem important to establish a minimum common set of parameters at European level and more flexible rules about financing.

At the more technical level of information, being able to evaluate the European-added value also means being able to rely on at least a minimum set of harmonised data inputs, and, in part, modelling outputs (including projections or forecasts). As discussed in the previous section, data availability and comparable information remains a problem. Furthermore, our knowledge on the underlying travel behaviour of individuals and companies and how this might differ across countries – an important element for modelling exercises – is still not adequate.

More generally the process of European integration has tended to strengthen the deliberative aspects of evaluation. Broadly, two views exist about what role the evaluation of transport proposals should have. These reflect different interpretations of the role and function of evaluation, and arguably, different world views also. One is simply that evaluation is a tool to assess value for money. A such, it has an auditing role, separate from and consciously independent of the decision process. An alternative view is that evaluation is very much part of a negotiation and deliberation process, through which socially desirable transport actions are gradually identified. It is not difficult to see that the process of European integration – itself a gradual process built on inter-governmental bargaining, negotiation and consensus building – would tend to underline that aspect of evaluation which is based on argumentation and communication.

This is also reflected in the increased role of stakeholders and the general public in the decision process. In this respect there are two types of challenges

faced: the first is that of involving those representing various interests in a discussion process with the objective of achieving a form of consensus or a compromise, at least on general principles; the second challenge concerns the process of consultation with users and the public more generally. These are challenges specific to the decision process, however they also influence evaluation practices as evaluation is called to support decision-making.

Stakeholders is a broad term used to describe those institutional actors or coalitions representing specific interests or having a stake in any decision. In transport they include the industry (across modes), operators, users' associations, cities, etc. Specific interests can also be represented as national interests, yet the term 'stakeholders' is not usually used to refer to Member States. The principle nevertheless applies. Stakeholder representation in evaluation is best achieved through their involvement at the early stage of defining the scope and terms of reference of an evaluation exercise and/or at the stage of discussing the available options on the basis of the assessment results.

The promotion of public debate on European matters is one of the key elements of the White Paper on European Governance which was recently published. A three-tier strategy is suggested for promoting this: (a) the provision of up-to-date on-line information on the preparation of policy through all the stages of decision-making; (b) a systematic dialogue with representatives of regional and local governments which are thought to more closely reflect the concerns of citizens than national governments; (c) seeking the pro-active involvement of civil society organisations in policy formulation – special mention is made of the social partners, i.e. trade unions and employer organisations.

In our representative democracies, direct citizen participation is often the exception than the rule. However there are occasions where it is advisable. In integrated technical assessment, focus groups or planning cells are methods often used for this purpose. In environmental impact assessment, public inquiries or hearings are institutionalised procedural steps in several countries.

More generally, how such calls challenge evaluation is with regard to communication and accountability. For evaluation to be accountable it must also be communicated in such a way that it is understood, including by non-experts. In as technocratic a field as transport evaluation, this is more difficult than might seem at first sight.

3.4 A multiplicity of methods

> "Today the policy analyst has available a multitude of assessment methods with differing underlying assumptions and a wide range of criteria for weighing and

> aggregating impacts and presenting results. No clear guidelines exist to tell which assessment method suits best which decision-making situation or how to integrate the results of the different methods" (Terms of Reference for TRANS-TALK, Additional Information Document, 1999).

There are several methods or tools that can be used for evaluation, separately or in conjunction. They range from survey and expert interviews for data collection to statistical analysis or modelling for data analysis or formal assessment techniques list cost-benefit and multi-criteria analyses. Evaluation methods can be classified according to four dimensions:

1. The analytical framework from which they emerged and within which they mostly operate;

2. Their suitability for policy, programme or project evaluation.

3. Their suitability for different phases of policy analysis (ex-ante, monitoring or ex-post);

4. The extent to which they rely on statistical or mathematical methods or models, or alternatively make use of 'softer' methods.

Existing classifications, like those proposed by government bodies or international organisations, tend to consider the second and third dimensions above only. In other words, they categorise tools according to their suitability for policy, programme or project evaluation and with regard to the timing of the evaluation exercise, but they do not discuss their analytical framework or the implications of the use made of statistics.

If classification schemes are to be good more than just as an inventory, they have to consider the two other dimensions mentioned above, namely the analytical framework of the method in question, and the extent of reliance on statistical or mathematical methods.

The analytical framework or metalevel of evaluation is not one which is today often dealt with explicitly. Either it is taken for granted that, for instance, cost-benefit analysis is rooted in welfare economics, or this is assumed not to be important either for the demand or the supply side of evaluation. This is the root of many flawed or unnecessary evaluations at either the project or programme level as each analytical framework entails different assumptions about epistemology and methodology; the complexity of problems; institutional settings or more generally the decision context, and, last but not least, welfare.

Let us consider as an example assumptions regarding welfare. It is well known, for instance, that cost benefit analysis makes some very strong assumptions about social welfare comprising the sum of individual welfare and being a reflection of the aggregate sum of the individual's 'willingness to pay', which is

clearly one and not necessarily the only possible definition of what social welfare comprises. Furthermore evaluation strategies differ as to the role they assign to the market regarding social welfare, but also with respect to the extent to which they are driven by considerations about efficiency, economy and effectiveness, or considerations about equity and ethics.

Several problems can arise by failing to be clear about the underlying assumptions of analytical frameworks used in evaluation. One frequent result is the subsequent failure to be clear about the objectives of evaluation exercises. This is a typical problem in multi-governance settings, like the European Union: the elaboration of policy programmes, and, subsequently, of evaluation tenders, represents itself a complex and often ill-structured process of bringing different and often contradictory objectives under the same hat. Another possible outcome is serious misunderstandings between those who commission and those who are assigned the task of evaluation.

Perhaps the most serious problem is the failure to understand or rightly interpret the results of an evaluation exercise and, in the case of a multi-tier evaluation exercise, to integrate the various results. Considering that policy analysis often necessitates such multi-tier evaluation exercises, it would seem that what often comes under the diagnosis of 'lack of co-ordination' is not simply a problem of organisation, but fundamentally also a problem in establishing a basis for 'reflective conversation'. The latter is only possible if at least a minimum level of common understanding is available.

The second dimension that is often not given adequate consideration concerns the implications of using specific tools, in particular statistical methods, for the object of study. There are two aspects to this. The first is that a minimum and up-to-date knowledge of statistics is necessary if these are applied. There are several examples which show that evaluation can go wrong due to the lack of statistical knowledge among those who apply it. Statistics is itself a science and in that relies on a set of assumptions. The evaluator needs to be aware of these prior to specifying an evaluation design.

The second has to do with the choice among relevant assumptions, data and methods of analysis. In order to be able to specify hypotheses and assumptions, it is important to have a certain understanding of the object under study. For policy-relevant issues, whether at the project or the programme level, this, in turn, requires a comprehensive knowledge of the policy in question, the decision-making context but also the role of policies as sets of interventions in the world of social relations.

The same arguments can be made also for tools like cost-benefit or multi-criteria analysis. The methodological development from cost-benefit (CBA) to social cost-benefit (SCBA) to multi-criteria (MCA) analyses represents a series of incremental attempts to overcome problems that are not solely theoretical but

also carry serious practical implications. Thus one major problem with cost benefit analysis is its theoretical reliance on market values and by extension shadow prices; the more flexible design, on the other hand, of multi-criteria analysis, can increase the likelihood of double counting. Both methods face problems with the specification of weights to apply to different criteria, albeit in different ways: cost-benefit analysis in view of the difficulties involved in measuring reliably the 'willingness to pay' of citizens; multi-criteria analysis in adopting a 'subjectivist' approach to this and relying on the decision-maker or a round of experts to determine how important any particular type of good or impact is for social welfare.

There are indeed today a multitude of assessment methods that can be used to carry out evaluations and arrive at policy-relevant results. How does one go about choosing the most appropriate? The above exposition has hopefully shown that this choice must be guided first, by a good understanding of the problem at hand including its context, and second, by a comprehensive knowledge of the methods and their analytical frameworks. The TRANS-TALK guidelines for improving evaluation in transport provide a conceptual map that can assist this selection.

3.5 Technical challenges

One of the questions posed to TRANS-TALK was whether existing methodologies are fundamentally constrained or whether they can be extended to address new challenges posed for evaluation. Some of these challenges like how to best link projects and programmes in evaluation or the involvement of stakeholders in the evaluation process were addressed in previous sections.

In this section we try to answer the above question for those issues that are more technical in nature:

- Transport and economic development;
- Network effects and European-added value;
- Time, term and uncertainty;
- Welfare basis of evaluation;
- Environmental valuation;

Scientific inquiry has progressed with regard to all of the above questions, opening new arenas for further research or offering clear recommendations on how to proceed with application in evaluation. Deliverable 5 of TRANS-TALK summarises some of these advances with reference to the papers presented at the 2nd TRANS-TALK workshop organised in Brussels in November 2000.

Some open questions remain. As is often the case with technical questions, the answers are to be sought in concepts and the definitions of terms. Disagreement or lack of clarity at this level has therefore to be resolved in advance. The following could be noted:

1. The degree and way in which transport is linked to economic development is still one unresolved issue in academic debate, with evidence pointing in either direction. It would seem that the closeness of association is heavily dependent on the area and network under study and their degree of development. How transport (policy) influences economic development must likewise be differentially assessed. This needs to be taken into account in modelling work.

2. Network effects and European added value. In a wider European context that is internally strongly differentiated (consider for one the North / South or West / East divides across Europe but also in several countries) it is important to examine transport developments both from the local and the global or wider perspective. Doing this brings into focus network effects that might have a considerable influence on the outcome and outputs of policies, and raises the relevance of the European added value. In particular the TEN-T projects can often not be appreciated or correctly assessed unless the European added value is taken seriously into account and operationalised accordingly in evaluation. Assessing TEN-T projects from merely a national perspective (even if from a plurality of them) does not suffice to capture the European added value or the adequate share of European financing.

3. Related to the above is also the question of how to deal with uncertainty in evaluation. This is particularly a relevant issue with regard to transport investments that are phased over several years, but also for policy measures like pricing that are likely to result in changes in individual or group behaviour that are not foreseen by static simulation modelling. The use of scenarios would appear particularly relevant for dealing with this problem, but also important is the identification of systemic and other inter-relationships between different effects.

4. The emergence of more complex evaluation problems is contributing to the re-emergence of older debates about the welfare basis of evaluation. Cost-benefit analysis assumes that it is possible for all costs and benefits to be valued using the same measuring rod (for instance, money) which implies, ultimately, that all types of costs are interchangeable. Several scholars and policy-makers see no problem with the theme of interchange or substitution between costs, and instead consider the fundamental issue to be that of measurement: for instance, how are environmental costs to be valued?

Those scholars operating from within the strong sustainability framework and often also representatives of environmental movements, argue against the valuation of some (or all) environmental effects with reference to the intrinsic or substantial (or even life-supporting) role some environmental sources have (hence the notion of critical natural capital). Time and uncertainty, in particular with regard to intra- and intergenerational effects, raise similar problems also with regard to equity and the question of justice.

With regard to the question of the capability of existing methodologies, there is no methodology that is fundamentally constrained. However and as discussed in the previous section, different methodologies are more useful in some contexts and for the analysis of some problems than of others. The combination of methods in a multi-disciplinary or inter-disciplinary framework offers more chances of success than the revision of standard methodologies in such a way that they lose their robustness.

3.6 Towards a better integration of political and technical perspectives

Transport evaluation has till now been mainly technocratic in focus. As a result there is a weak link between the community of transport professionals who undertake the analysis and decision-makers or their policy advisers, which can (and often does) result in communication failure.

Policy-makers and politicians often complain that (a) they do not get what they asked for; (b) they do not get the input at the right time; or (c) that they do not get it in a format which is accessible to non-experts and thus policy-relevant. On the other hand, transport professionals worry about political intervention that might jeopardise their autonomy; or they complain about the tendency of politicians to expect simple answers to complex problems.

When seeking to better integrate political and technical perspectives in evaluation it is important to recognise that politicians, on the one hand, and scientists or professionals, on the other, have distinct functions which inevitably lead them to have different expectations from evaluations. A politician called to make a decision on infrastructure investment carries the responsibility for the decision and is accountable vis-à-vis his/her executive and/or legislative as well as the electorate. He or she is expected also to be able to defend the decision to the various stakeholders and possibly to independent experts. The audience of the politician is not an expert audience, hence the call for evaluation results which can be processed and understood. This is a very different world from that of a scientist, the public of which is expert and more critical and who is expected to approach any subject in a comprehensive way that tends to underline the latter's complexity rather than simplicity.

The role of the evaluator is in that problematic because it represents (or ought to represent) an intermediary position between the politician who decides and presents and the scientist who studies and criticises. Even though it is often scientists who act as evaluators, the skills that are most important for an evaluator are not necessarily those most important for a scientist. Whereas for a scientist what counts is specialisation, including technical or academic skills, for an evaluator it is often more helpful to be a generalist (at least in some subjects) and to display an ability to communicate findings in user-friendly language. Not all scientists make good evaluators (or vice-versa) just like not all teachers make good researchers (or vice-versa). Important for evaluation perceived as a process is that all of the above functional roles are represented through the setting up of an appropriate team.

On the subject of autonomy, it would be naive to deny the political interest in evaluation. Again however this is, in itself, not a problem. The political interest in evaluation is only a problem if it is one-sided, reflecting vested interests or not accountable. This is also why evaluation ought not to be the monopoly of government or dependent on governmental or state funding. It should also not be primarily or solely carried out internally.

In fact both internal and external evaluations are important and necessary and nothing would speak against having both, provided they are balanced accordingly. Over-reliance on internal evaluation is likely to lead to lack of transparency and a decrease of accountability. External evaluation is more likely to be autonomous and contribute to institutional learning. More important perhaps in either case, is that the results of evaluation studies – whether internal or external – are made public and/or submitted to consultation. More generally, the diversification of (transport) expertise must be promoted and encouraged as it is ultimately what ensures that participatory models of decision-making work also in terms of output.

Finally there are two other factors that are very important for promoting the better integration of political and technical perspectives in evaluation. The first is realism and applies to both the demand and supply sides. Evaluation is a science of interpretation. No method can deliver objective and uncontested findings. It is neither a substitute for decision, nor a substitute for deliberation.

The second factor is organisational and reflects the fact that evaluation is a process. Evaluation begins already with the definition of an evaluation plan and ends with the processing of results to design or change policy. Just as it is important for evaluation teams to reflect on the production and dissemination of knowledge, recognising that these are two different functions that might also have to be separated at the personal level, commissioning agencies need to recognise that evaluation does not end with the commissioning of a study but that it is important to install procedures that ensure that the evaluation results

are monitored and, once available, reach those that need them in the most appropriate format.

4 Recommendations

The TRANS-TALK thematic network has established a framework for thinking about transport evaluation in a comprehensive manner, across national borders, i.e. with reference to common or shared objectives at European level, and in a way sensitive to the need for deliberation across partly or potentially conflicting interests.

On the basis of the knowledge gathered and experiences made with the TRANS-TALK thematic network, the following recommendations could be advanced. To exemplify the recommendations, reference is made, where possible, to the new White Paper on European transport policy.

4.1 Frameworks of evaluation

4.1.1 The policy level

The new White Paper on the European transport policy recently published is interesting in several ways, and not least for recognising the importance of monitoring and evaluation – a first progress report to the European Parliament is due in 2005.

At this broad policy level, there are two types of evaluation that are relevant:

a) output evaluation (at regular intervals) to examine or check whether the action points or measures announced have also been implemented;

b) outcome evaluation at the short-term (for instance 2005), the medium term (the end of the policy cycle to which the White Paper is referring, i.e. 2010) and the long-term (2020 and then 2030 when it is estimated that the policy initiatives which are hereby launched will have come to bear fruit).

Both types of evaluation are important, yet both have to be planned and implemented carefully.

With regard to output evaluation at policy level, once the measures and/or action points identified have been listed and clustered thematically (as in the White Paper on European transport policy and most master- or policy plans at national level), the following are important steps:

1. Specify for each of the proposed measures a detailed time plan for implementation taking into account the decision procedures that might be

involved. In the White Paper this is done in part (thus, for instance, with regard to the directive on the TEN-T it is said that a first revision will have been completed by 2002, the second by 2004), but not systematically (thus it is for instance not said when the process of re-thinking airport capacity and use in consultation with Member States and relevant stakeholders will start).

2. To the extent possible, re-organise the time plan in such a way as to make maximum use of possible synergies. This is of particular importance for those policy measures or initiatives that are still at the first consultation phase and largely relying on the momentum to be gained through expert meetings or voluntary agreements emerging bottom-up.

3. Once such detailed time plans exist for each measure, it is possible to determine the best milestones for monitoring purposes.

4. Where the measures under evaluation concern European-wide directives, it is important that it is recognised that the sole transposition of a directive in the national environment, while a first essential step, does not yet guarantee implementation. Thus the time frames for monitoring should take that into account.

Output evaluation says little about the success of the proposed measures in effecting the changes envisaged – in the case of the White Paper on European Transport policy with regard to reducing congestion (through effecting a modal shift in favour of rail, maritime and inland waterways), and dealing with environmental externalities (through the introduction of road pricing, the revenues of which shall be earmarked for subsidising infrastructure investment in environmentally-friendly modes).

Outcome evaluation at the policy level for the purpose of monitoring is best carried out at the aggregate level with reference to broad indicators, for instance,

- modal split (overall and spatially distributed at national and/or regional level and separately for freight and passenger as well as for short-distance and long-distance travel), or

- CO_2 emissions (globally, nationally, locally and for specific environmentally sensitive areas or modes),

among others,

- set against structural indicators specific to the transport sector (like traffic volumes per mode, globally, nationally and for short-distance and long-distance travel).

A detailed list of indicators for monitoring purposes would have to be elaborated through deliberation with Member States to guarantee coverage, the necessary level of scope, and detail as well as a common measurement and methodological basis.

The disadvantage of aggregate level monitoring based on indicators such as the above is that it, of course, does not provide precise information on the interrelationships between measures and the precise causal relations (within transport but also from factors external to the transport system). Thus, for instance, a possible reduction of environmental emissions might be the result of not alone a change in the modal split but also of a reduction in travel demand derived from a slowing of the economy. How much of the effect is, in turn, due to the change in the modal split, and how much to the reduction of travel demand due to lower economic growth, can only be established on the basis of time-series data over several years, or through the aggregation of results of more detailed studies at the programme or project level.

A combined monitoring system combining output- and outcome-specific indicators as described above is, despite its analytical limitations, a support tool that no decision-making level can or should do without, if policy evaluation is to be taken seriously.

Such a monitoring framework displays yet other advantages. It can contribute rather quickly to the harmonisation or integration of technical parameters (with regard to data collection or analysis). Perhaps more importantly, it has a contribution to make to consensus-building and co-ordination processes with regard to policy objectives or measures, in that it provides a common basis for the latter's assessment in time.

4.1.2 The programme level

A programme represents a set of projects that can be clustered together in view of their overall policy objectives (for instance, aiming at reducing the burden on the environment) and/or intervention mode (investment, regulation, economic measures, research, etc.) and/or scope of application (geographical, mode, etc.) and/or organisation structure (for instance, framework of implementation). Programmes are often defined at the outset by the policy-maker or they can be derived independently from the evaluator or researcher in the course of analysis.

At the programme level it is possible to carry out all three main types of evaluation, i.e. output, results and outcome oriented evaluation. Output evaluation provides information on implementation. The difference between outcome and results evaluation is one of degree. Whilst the former focuses on long-term indirect impacts of general validity or with reference to the policy

objectives, the latter focuses on the direct impacts on a specific target population in the short-term.

Whereas output evaluation regarding programmes represents theoretically an intrinsic part of policy design, the same is not true for evaluation that concentrates on results or outcomes. The latter need to be timed well to coincide with important milestones in implementation when the impacts or effects can be expected (whether short-term or long-term).

The level of detail of the evaluation of a programme's results or outcomes depends on the character of the programme. Some programmes are so multifarious that evaluation can only be carried out at the aggregate level, as proposed in the previous section (4.1.1) for policy. This would, for instance, be the case of the evaluation of the programme or policy package of measures that the Commission plans to implement to reduce the environmental externalities of transport as these range from road pricing to investment in environmentally friendly modes and the harmonisation of technical standards. Likewise with the policy package being proposed for improving the efficiency, organisation and sustainability of air transport. Other programmes are more clearly delineated, thus allowing for the application of more detailed evaluation techniques. For example, it is possible to carry out a strategic assessment that employs cost-benefit and multi-criteria analysis on a set of infrastructure projects comprising a network.

4.1.3 The project level

A project in transport is often taken to refer to an infrastructure project that is strictly delineated in terms of space and time. Evaluation practices in transport are most developed at this level and especially concerning appraisal, as we saw in section 3.2 of the previous chapter. To reiterate, we found that most countries use cost-benefit and multi-criteria analysis to assess transport infrastructure projects, and that the existing evaluation frameworks incorporate all direct transport costs as well as the most important external costs, albeit different valuation criteria or weights are applied to the latter. Here, the need for improvements concern primarily the European level and, related to this, the demand for harmonisation.

More specifically, it would be advisable to arrive at commonly agreed parameters with regard the assessment of projects declared of European interest (in the framework of the TEN-T) in order to correctly reflect the latter's European added value.

At the more aggregate level, a certain degree of harmonisation is called for with regard to the background assumptions to guide forecasting exercises. This

includes macro-economic assumptions, assumptions about the development of the transport sectors including costs as well as impacts.

4.2 The context of evaluation

4.2.1 The policy context: the black box of evaluation

In policy analysis it is often useful to refer to best-practice models or benchmarking exercises when planning and implementing policy reform. However, the transferability of best-practice models (from one country to another) often turns out more difficult than originally envisaged. The reason has to do with the context of policy-making which relates as much to procedures, including to political culture, as it does to contents. This is something that needs to be paid special attention in the evaluation of policy, both with regard to outputs and outcomes. Thus the evaluator must ask and document whether a policy was implemented and/or how successful it was, yet must also inquire into those factors that have facilitated or obstructed its implementation.

4.2.2 Beyond value-for-money towards deliberation

Complex decision-making procedures – like those characteristic of European level governance in between intergovernmental bargaining and integration – call for a new culture of evaluation that places more emphasis on deliberation, especially with regard to appraisal.

This need not strictly imply that it is the task of the evaluation team to organise a deliberation among relevant stakeholders (or with the public through planning cells or public inquiries). Deliberations of this type have a political function that often goes beyond the terms of reference of any technical evaluation study. However where such deliberations have taken place in the process of decision-making leading to a policy proposal and the commissioning of an evaluation study, it is important for the evaluation team to be aware of these procedures, their results and implications for the evaluation. For instance, it could be that the evaluation study was commissioned as providing input to an ongoing deliberation process among relevant stakeholders, in which case it will be important for the evaluation team to meet with the various stakeholders and discuss their expectations with regard to the evaluation study.

More generally, it is important to carry out and communicate evaluation studies in such a way that the procedures followed may be repeated, thus followed, by

external reviewers. This increases both the accountability and transparency of evaluation. Making the latter open for critical review is a major step towards establishing a deliberation culture in evaluation.

4.3 Methods of evaluation

4.3.1 Everything goes? ... Not quite

There are several methods that can be employed in transport evaluation, be it with regard to data collection, analysis or formal assessment. The multiplicity of methods can be mistaken to imply that everything goes, more specifically that either any (desirable) result is possible to deliver if the choice of methods is carefully thought through, or that differences in evaluation results are methodologically grounded.

Even though the choice of methodology can and does influence results, relativism is however not the conclusion to be drawn from this. Robust findings will remain even when using different methodologies. Otherwise, there are effects that might not be tapped by certain approaches, not because these do not exist, but rather because they did not form part of the subject of inquiry.

When thus comparing the results of different evaluation studies on the same policy, programme or project, it is important to establish first, whether the subject and scope of inquiry was defined in the same or similar fashion; second, whether the methods used for data collection, analysis and formal assessment were of the same family or type, and third, how similar or different the background assumptions, impact functions etc. were.

Needless to say, quality standards apply with regard to the use of methods and these can usually be read in standard textbooks or literature on the subject. Failure to apply these standards can indeed lead to distorted evaluation results, yet these should be easy to identify and assess.

More generally, the choice of an evaluation method should be guided by a good understanding of the subject of study and a thorough knowledge of the methods to be used.

4.3.2 Time and the use of scenarios

Assuming data availability, ex-post evaluation is in theory easier than ex-ante evaluation or appraisal. In practice there is however little difference, not only

because of the data quality but also because policies, programmes or projects are not always or even most frequently processes with clear beginnings and ends. Taking this argument a bit further, one could even say that there is always an element of appraisal in ex-post evaluation and of ex-post evaluation in appraisal.

Both types of evaluations, but especially appraisal, have to rely on a set of assumptions about the way the various policy measures, impacts or target impact groups are interrelated, and the direction of cause and effect at different levels. For some of these interrelationships there exists clear objective evidence that can be used. However unlike what is often thought this is the exception rather than the rule. The objective evidence, where available, is often particular to specific conditions or time periods that make generalisations problematic. Also often in transport, the evidence available is statistical in character and the latter is based on probability theory.

It is for this reason that it is important in appraisal to work with scenarios that define various options or visions of the future or the transport system. Besides making transparent how conclusions are derived, thus contributing to policy deliberation, the use of scenarios allows more flexibility in evaluation thus contributes to the latter's and to policy's results.

In transport scenarios should be developed with regard to:

- Macro-economic and/or political developments;
- Sector trends, including technological options;
- Policy packages.

Most policy plans and appraisal studies operate either with only one or two types of scenarios or with scenarios defined in such a way that makes comparison difficult.

Thus, for instance, the new White Paper on European Transport Policy relies on only one macro-economic scenario (that in turn rests primarily on rather optimistic assumptions regarding GDP growth) for the year 2010, which subsequently is used to determine sector trends (for the different modes). Against these, three policy packages, which differ with regard to their degree of inclusiveness of measures other than road pricing, are simulated. The definition of policy packages in a cumulative fashion allows to account for the progressive impact of additional measures other than road pricing (in addition to highlighting the importance of the latter).

Methodologically this approach is legitimate and has clearly the advantage of being simple and easy to communicate. However it cannot substitute for a more comprehensive appraisal or forecasts of transport demand, modal split or

emissions for the year 2010 which would consider additional scenarios. What is likely to happen, for instance, if road pricing is not implemented across Europe but only in specific regions (with or without additional measures)? Or what happens if road pricing is indeed implemented but there instead follows a period of recession or low economic growth, rather than a period of high economic growth? The underlying assumptions about impact functions need not change for these other simulations. Yet what such simulations would deliver is a type of sensitivity analysis that would allow to establish with more certainty which policy measures are absolutely indispensable and which are less essential (at least till 2010). This in turn could then be fed into output evaluation at policy or programme level and help better plan the policy or implementation process.

4.4 Technical challenges – new research priorities

Complex evaluation problems require often new knowledge and the latter can only be obtained through science, research and development. Evaluation and research are two overlapping, yet distinct, arenas. It is important for evaluation to remain up-to-date with developments in the RTD field. Likewise, it is important for research to think always of applications, including in evaluation.

The review of contemporary evaluation challenges in transport has revealed a series of subjects for which research is needed. We can talk of two types of research priorities:

For some areas, the type of research which is currently needed is more of a technical, development or demonstration type. For instance, we now know that the relation between transport and economic development is complex and there is enough empirical case study material that exemplifies these complexities. To advance the state-of-the-art in this field, what is now needed is research that attempts to specify these complexities in modelling work.

In other areas on the other hand, there is a lack of either empirical information or of conceptual elaboration. Empirical information is, for instance, lacking on the subject of the transport mobility of individuals, especially on the dynamic aspects of mobility when faced with new options (new routes, new infrastructure, new technology, etc.). Conceptual work, on the other hand, is needed to advance our understanding regarding the meaning of an environmentally sensitive area and the valuing of environmental effects, or the interrelationships between transport and non-transport policies.

4.5 Towards a better integration of technical and political perspectives

The better integration of technical and political perspectives in evaluation requires changes both from the technical and from the political side. More realism and a greater willingness to co-operate are minimum requirements. Beyond this, the different functions of evaluation, on the one hand, and of politics on the other, have to be recognised and respected.

5 Annex A. Summary of Workshop Proceedings

5.1 Policy and project evaluation: context, theory and methods

5.1.1 The agenda

The first TRANS-TALK workshop addressed the context of project and policy evaluation and how this influences the definition of the problem, the evaluation methods selected and their application. Five sessions with commissioned papers were organised over three days:

The first session 'The Context of Policy-Making' looked at decision-making processes for the transport and environment sectors at Union level in general and in view of enlargement;

The second session 'Analytical Frameworks and Methods' provided an overview of evaluation frameworks and methods in use generally, and in the fields of transport and environment in particular;

The third session 'Problem Definition, Interests and Ideas' was a brainstorming session offering participants the opportunity to reflect in groups on how policy problems are defined and, in turn, how this influences evaluation designs;

The fourth session 'Evaluation Methods Compared' compared the evaluation methods in use by international organisations and for different sectors;

Finally, the fifth session 'The National Perspectives' examined the national perspectives in transport evaluation.

5.1.2 Summary of presentations

The Context of Policy Evaluation: Institutions and Decision-Making Processes

The Policy Making Process in the European Union
Francis McGowan, Sussex European Institute, University of Sussex

The paper examined the main aspects of the EU policy making process, particularly as it affects the areas of energy and transport policy.

In outlining the historical development of the Union and of its institutions, the author noted that the tension between intergovernmentalism and supranationalism, and in this connection, between the European and the national levels of decision-making, are reflected in the original structure of the Union. The EU displays a number of distinctive features which make it not easily classifiable as either a nation-state or an international organisation. The author identified three such features: first, that the EU is a polity, but an incomplete one; second, that it has a propensity towards regulation, and third that it is based on multi-level policy-making.

In order to understand whether there has been (or will be) a convergence in policies, including in the transport and energy fields, it is necessary to understand the institutional interactions. This applies both to interaction between European institutions (in particular between the Commission, Council, Parliament and Court of Justice) and interactions within these institutions. Other relevant EU institutional characteristics are: the character of lobbying at European level with multiple paths of interest representation; multi-level governance and subsidiarity, i.e. the complex interface between European, national, regional and local levels of government, partly lacking an explicit or agreed upon 'division of labour' or modes of co-ordination; and the Europeanisation of policy and implementation gaps.

With regard to policy change, the author distinguished between three logics of policy change, namely systemic, partisan and sectoral, and argued that the systemic logic has dominated EU policy change till now; in turn this underlies the importance of study of national policies and of alliances between nation states on policy agendas.

The EU Enlargement and its Impacts on European Policies
Gerhard Rambow

This paper reviewed the process of enlargement and its impacts on European policies. It addressed the issue of why the Eastern Enlargement poses a bigger challenge for the Union than the previous enlargements.

Some of the factors that will be greatly affected by the Eastern enlargement are:

- the functioning of the Union, with regard to its own institutions;
- the principle of subsidiarity, considering the greater diversity to be introduced with the new members states;
- the process of formulation of policy as the interests and ideas of the new member states will have to be incorporated; and
- the process of execution or implementation of policy -- there are already implementation gaps with regard to policy and these might increase with enlargement. The author argued that the existing flexibility clause with regard to implementation is not enough to accommodate for diversity.

The paper outlined the possible implications of the Union's enlargement for redistributive policies (specifically the budget of the Structural Funds); agricultural policy; cohesion policy and transport policy.

The Europeanisation of policies

Recent Developments in the Common Transport Policy
Vicenc Pedret Cusco, EC DG-TREN

This paper looked at the challenges of meeting the Kyoto targets as well as the increase in traffic expected with the enlargement of the EU. In the new Community Transport Policy, emphasis will be placed on the better use of capacity, mainly on rail, inland waterways and short sea shipping; intermodality; interoperability; network interconnectivity, and bottlenecks. Other priorities of the Common Transport Policy are the better integration of environmental aspects in infrastructure planning; internal and external cost coverage; and the integration of transport policy concerns in other policies.

Enlarging EU Environmental Policy; The challenges of flexibility and integration
Ingmar von Homeyer, Centre for International and European Environmental Research

The upcoming enlargement of the European Union has always been described as a major challenge for the EU environmental policy. This paper looked at the question of why failure of the Candidate Countries to fully comply with European environmental legislation at the date of accession (assuming that accession takes place in the medium term) should pose a major challenge to European environmental policy. In trying to answer this question, the paper assessed the various elements of the EU Environmental policy, which are:

- The leader-laggard dynamics in the making of the European environmental policy;
- The implementation deficit characteristic of EU environmental legislation.
- Political efforts by Member States and the European Commission to reform EU environmental policy. The integration of environmental concerns into other sectoral policies is probably the most important and promising element in the ongoing efforts to increase the effectiveness of EU Environmental Policy.

Analytical frameworks, principles and methods of evaluation

Analytical frameworks for Policy and Project evaluation – From Welfare Economics and Public choice to Management approaches
Wayne Parsons, Queen Mary and Westfield College, University of London

The paper examined the varieties of frames used in policy evaluation, showed how these embody different 'assumptive worlds' and examined the implications of the latter for theory and practice of evaluation. It distinguished at least eight analytical frameworks for evaluation: neo-classical economics, experimentalism, managerialism, public choice, pragmatism, interpretivism, evaluation through the price system, and critical realism. These frameworks apply equally to policy analysis or analysis for policy as they do to evaluation.

Each of these frameworks entail different assumptions about: epistemology and methodology; the complexity of the problems addressed; institutional settings or more generally the decision context; and welfare.

The paper also looked at the question of integration in evaluation. The author argued that this can be done through a multi-methodological strategy which distinguishes between first and second order evaluation. First order evaluation moves from programme verification to situational validation. Second order evaluation moves from systems discourse to the ideological discourse and focuses on the macro level using hermeneutic-interpretivist approaches. Better integration can be achieved through more frequent and better use of scenarios in evaluation.

Evaluation of Projects and Programmes: Principles and examples
Frank A. Haight, Univ. of California Irvine

This paper looked at the development of project evaluation and uses examples from traffic safety which reflect the classification and vocabulary used by the United States Department of Transportation. More specifically it looked at the issues important for project evaluation, the principles of evaluation and the types of evaluation.

With respect to project evaluation, the paper addressed firstly, the issue of the need to be clear about the parameters of the problem to be evaluated, and secondly, the need to be aware that not all problems deserve evaluation.

The principles of evaluation outlined in the paper are that: the evaluation plan must be specified in advance; the evaluator should be unbiased; and that statistical tools need to be used carefully and wisely, recognising that human judgement is part of the process.

With respect to the types of evaluation, the paper distinguished between evaluation with reference to the project timing (planning, implementation and evaluation) and with reference to the methods used (quantitative or qualitative).

Methods of Transport Projects Evaluation – From Cost-Benefit to Multi-Criteria and Decision Framework
Michel Beuthe, FUCAM

The paper reviewed the problems faced by evaluation/assessment with reference to cost-benefit analysis, multi-criteria and framework analyses.

For CBA the problem lies in the understanding and definition of market values, their measurement in situation of non-competition and non-perfect competition, the valuing of impacts which do not have a market, the choice of the discount rate, national differentiation, income distribution and risk assessment.

The main problem faced by MCA, on the other hand, is that of 'easy arbitrary assessment' as the main reference of valuing the project is the public decision-maker. Other problems with MCA are that of double counting effects due to inclusion of qualitative or political criteria; the shift from differential to unitary scale of measurement and its inability to assess the projects' worth from the budgetary perspective.

The author argued that it is necessary to consider the application of evaluation methods as part of a larger decision framework.

Criteria for Evaluation towards Sustainability
Klaus Rennings, Sigruid Weinreich, ZEW

The paper reviewed the ecological criteria for sustainable mobility as set out by the OECD, defined the principles for the sustainable evolution of transport systems and the application of the sustainability criteria in cost-benefit analysis.

The author defined the environmentally sustainable transport as representing a process, or a path to be followed. Sustainability is not a fixed ideal, but an evolutionary process of improving the management of systems, through improved understanding and knowledge. The process is non-deterministic with the end point not known in advance. For this path nine sustainable transportation principles – the so called Vancouver principles – were developed based on human perspectives: access, equity, individual and community responsibility, health and safety, education and public participation, integrated planning, land and resource use, pollution prevention, economic well-being.

For integrating sustainability aspects a modern concept of cost-benefit analysis was introduced in the Second Assessment Report of the Intergovernmental Panel on Climate Change. CBA provides an analytical framework that seeks to compare the consequences of alternative policy actions on a quantitative rather

than a qualitative basis. The internal and most of external costs and benefits can be analysed by using the traditional CBA which requires that all costs and benefits are expressed in a common monetary unit. If the external cost can be calculated and internalised by using the damage cost approach no sustainability criteria will be applied.

Concerning the climate change issue or the valuation of impacts on nature and landscape (biodiversity), costs and benefits have been estimated in several studies but the results are highly uncertain and controversial. Thus, other techniques such as cost-effectiveness analysis or multicriteria analysis may be used for the inclusion of new transport-specific sustainability criteria.

Perspectives of International / Professional Organisations on Evaluation

Transport Project Appraisal in the World Bank
Ken Gwilliam, World Bank

In the World Bank the main purpose of economic analysis is 'to help design and select projects that contribute to the welfare of a country'. This paper outlined the evaluation framework used by the World Bank for appraising projects, more specifically transport projects. It specified how economic analysis is much broader than traditional cost-benefit analysis. The ten questions that an economic analysis in the World Bank looks at are:

1. What is the objective of the project;
2. What will happen if the project is implemented;
3. Is the project the best alternative;
4. Are there any separable components and how good are they separately;
5. Who are the winners and losers;
6. Is the project financially sustainable;
7. What is the project's fiscal impact;
8. What is the project's environmental impacts;
9. Is the project worthwhile;
10. Is it a risky project?

Much emphasis is placed on the institutional and regulatory environment into which the bank is lending. Project lending which contributes to improvement of that environment is thus of particular interest. Judging the extent to which that is being achieved depends on some of the high level priorities which the Board of the Bank have adopted. These include the avoidance of environmentally damaging investments, the widespread distribution of the benefits of projects throughout the national recipient community and the avoidance of uncompensated losses by virtue of spatial or occupational displacement resulting from a project. Strict standards are applied both to the environmental design of projects, which have to have formal environmental clearance before they are submitted for Board approval, and to the resettlement and involuntary

employment severance. The requirement to attend to distributional aspects is increasing as the Bank concentrates further on its poverty reduction objective.

All of that contributes to a broadening of the format of project appraisal, and in a sense to attenuate the more conventional use of any strict ranking procedures based on traditional cost / benefit ratios. But in a different sense conventional economic evaluation has remained central to the appraisal process as there remains an extremely strong instinct to avoid 'white elephants', particularly those emanating from political glory seeking. The process through which that is pursued is essentially one of progressive refinement of the understanding of national priorities.

Evaluation Methods in Energy – The EIB experience
Juan Alario, EIB

In the EIB the main objective as set out in the EC Treaty is to facilitate the financing of projects contributing to developing less developed regions and to the progressive establishment of the common market. Projects of common interest for several member states are given priority. This paper specified the three stages of evaluation in EIB, namely, appraisal(ex-ante evaluation), monitoring and ex post evaluation and the criteria used at each of these stages.

In the appraisal stage the contribution of a project to Bank's objectives, including its economic, financial and technical viability, as well as its environmental sustainability, are evaluated. More specifically, project appraisal in the EIB covers the following aspects: context, economic interest, eligibility with regard to EIB objectives, market analysis, particular technical issues, environment, procurement, investment costs, implementation, operation and financial and economic profitability. The main decisive criteria in the project appraisal are the economic justification of the project, which incorporates technical and environmental criteria, and its contribution to the Bank's objectives.

In the monitoring phase, during project implementation, the financial and physical progress of projects is followed until they are completed. Finally, in the post-evaluation phase (usually at least 18 months after the project has started running), the project performance and its contribution to EIB objectives are re-evaluated.

Appraisal issues for Trans-European Projects
John Nellthorp, ITS Leeds

This paper gave an overview of the current appraisal methods in the accession countries and the EU15. It highlighted the way methods are adapted to address policy concerns other than the narrow economic efficiency of the transport system but also aimed to draw out some difficulties with current approaches. It

considered the treatment of environmental impacts, effects on economic development, employment and land use, and effects on accessibility, social cohesion and regional policy.

In order to help prioritise the various individual projects, and to make the case for funding by the EU through ISPA and by the international financial institutions, a common methodology for project appraisal was introduced through TINA. In brief, the main outputs are: cost-benefit analysis - covering effects on transport system efficiency and safety; environmental impacts; policy impacts beyond the transport system; and financial implications for transport providers. The focus of the appraisal is directly related to the main policy concern of the funding bodies - i.e. obtaining the best value for money in socio-economic terms for the countries concerned.

Evaluation Methods for Roads
Ansgar Kauf, IRF

This paper outlined the views of the International road Federation (IRF) for the evaluation methods to be used for Roads. It assessed the use of cost-benefit analyses, multi-criteria analyses and framework analyses for appraising road projects. More specifically, it reviewed the use of social cost-benefit analysis for project and policy evaluation methods, the financial appraisal of projects and programmes and the strategic assessment and decision-making processes in evaluation. For cost-benefit analysis the relevant criteria to be used for roads are, direct financial costs and benefits to users; financial viability of the project; overall economic and social worth; safety and health of population; environmental considerations; chances of realisation of the project and the identification of obstacles to its implementation.

The paper concluded that it would be most beneficial to use evaluation methods based on a financially quantified, integrated socio-economic cost-benefit analysis and to have democratic decision making processes.

External Costs of Transport
Gunther Ellwanger, UIC

This paper outlines the results of a study commissioned by the UIC to estimate the external costs of transport. External costs are causing wrong market signals leading to significant inefficiencies. Railway systems are the environmentally most adapted transport means, but are still negatively affected by non-internalisation of external costs. An internalisation of external costs towards fair and efficient prices between transport means is essential in order to develop sustainable transport solutions. In order to implement such a policy, a sound empirical basis of external costs is imperative.

National Perspectives on Transport Evaluation

A roundtable with national representatives concluded the following regarding transport evaluation methods at the national level:

In most countries transport evaluation focuses on infrastructure planning and assessment and in this connection ex-ante programme or project evaluation. In some countries a multi-modal perspective is taken, i.e. masterplans for infrastructure cover both rail and road; in some a modal perspective is taken, i.e. road and rail administrations operate separately. In no country is there ex-post evaluation of infrastructure programmes.

In most European countries no major data or model constraints were identified. The exception was East European countries.

Environmental issues are considered in evaluation as part of project assessment and/or as part of programme assessment. Equity, i.e. distributional issues are less often considered with the exception of accessibility.

There would appear to be no systematic practices for ex-post project or programme evaluation or for policy evaluation (both ex-ante and ex-post). In the latter connection, the evaluation could be said to be more of a political nature or relying on the commissioning of expertise, whereby procedures are not very transparent. Only in a few cases (for instance Netherlands) are elements relating to the more general transport policy environment (like regulations) considered either for establishing reference scenarios or for 'correcting' assessments.

Regions are responsible for regional planning and public transport, yet they are not obliged to use the methods of evaluations used at the national level. The interfaces between the national and regional levels are unclear; so are those between the national and European levels.

There would appear to exist no harmonised criteria about prioritisation, phasing and forecasts.

5.2 Projects, programmes, policies: evaluation needs and capabilities

5.2.1 The agenda

The second workshop was concerned with outstanding problems in project evaluation methodologies and the links between evaluation at the project and the programme level of assessment.

Contributions were invited that addressed one or more of the themes below from the perspective of whether existing methodologies are fundamentally constrained, or whether they can be extended to address new challenges posed for evaluation.

Transport and economic development
How to identify and evaluate the impacts that transport investment or policy changes may have on patterns of economic development at European and other levels, exposing spatial redistribution of activity, and avoiding, for example, double-counting.

Social exclusion and distribution/equity
What aspects, if any, of social exclusion are legitimately seen as consequences of transport sector decisions? At what physical scale of analysis does it become appropriate to measure and evaluate exclusion? What indicators of social and spatial equity are feasible and appropriate to use in project and programme evaluation?

Involving stakeholders in the evaluation process
How can evaluation become more open, transparent and accountable to the final users of transport systems? What forms of participation would be appropriate to involve stakeholders in the evaluation process? What evaluation methodologies are consistent with stakeholder involvement?

Environmental evaluation – CVM/SP
How close are we to being able to include all environmental impacts in cost-benefit analysis? How reliable are the estimates of impact value and what estimation procedures are preferred? Are strategic environmental impact assessments consistent with project-level evaluation procedures?

Network effects and European value added
What impacts of major schemes are potentially overlooked if individual member states simply evaluate them at a national level? How, in practice, can network effects at the European level be modelled and evaluated? To what broader EU policy goals might such schemes contribute that are not captured in typical evaluations and how can these contributions be evaluated?

Welfare basis of evaluation
What are the unstated but significant implications of using cost-benefit analysis or multi-criteria analysis as a basis for project evaluation? Does either truly have the capacity to fulfil the socio-political role expected of it? Are these ways of evaluating acceptable at programme of policy levels?

Time, term and uncertainty
Do current evaluation procedures adequately address uncertainty? Are the methods used to aggregate impacts over time acceptable in the context of strategic transport planning?

In addition, contributions were invited for two parallel sessions that explored, respectively, evaluation of the costs and benefits of alternative pricing policies, and evaluation in the context of Trans-European Networks.

5.2.2 Summary of presentations

Strategic Planning and Assessment at National and European Levels

Strategic Transport Planning and Evaluation: The Scandinavian Experience
Henning Lauridsen, Institute of Transport Economics, Norway

Ex post evaluations in Norway and Sweden have revealed considerable shortcomings with regard to the planning process and evaluation methods at the national level. The evaluations reveal a number of shortcomings in the assessment methodology and tools, even where considerable efforts have been made to extend previous methods and develop new ones. The paper discussed whether the current planning methods and planning tools address strategic planning needs. Recent studies have shown, for instance, that politicians only rarely consider and use the information on strategic options entailed in masterplans.

A key question is whether the existing planning and assessment methods can indeed handle prioritising across sub-sectors and if they can take into account types of interventions other than infrastructure development. A second key question is whether current evaluation methods are applicable at more than one planning level, e.g. the national level, the corridor level, or the regional level. A third question concerns the applicability of methods for both long-term plans, say with a time perspective of more than 10 years, and for action plans with a shorter time perspective.

Impact Assessment of Strategic Road Management and Development Plan of Finnish National Road Administration
Eeva Linkama, Finnish National Road Administration

The Guidelines for Road Management and Development 2015 constitute the strategic long-term plan on road management and development in Finland. These introduced some new elements in the strategic planning in Finland regarding road. The main new element concerns the integration of impact assessment into planning. The plan is based on the impact assessment of three alternatives and their comparison. qualitative rather than quantitative analysis was given priority and emphasis was put on measuring performance with regard to specific goals which were themselves prioritised. Participation in preparing the plan included other government organisations in the transport sector, national interest groups, the road districts and regional interest groups, as well as other clients and partners, based on the principle of representation. Once finalised the plan was submitted to an external audit. The auditors concluded that despite some problems remaining, overall this new way of planning had been a useful learning process and recommended its use at regional level as well.

Several recommendations were advanced for improving the planning process. Thus it was found that more emphasis should be placed on defining the goals of road management and development. Many of the goals were too vaguely defined, leaving too much room for interpretation. Furthermore it was suggested that in the future the overall guiding goals are reduced in number. Qualitative analysis in strategic planning should be prioritised, however where quantitative analysis is unavoidable it would be necessary to ensure the adequacy of the baseline information. In this connection more information on travel patterns (also with regard to leisure and non-work) is necessary.

SEA and its Relationship to Transportation Projects
Paul Tomlinson, Transport Research Laboratory, UK

The paper provided an introduction to SEA as well as to the current proposed requirements of the draft SEA Directive. Perhaps more than any other development sector, transport is well positioned to lead the application of SEA given the structured approach to the delivery of transport measures that generally exists. The emerging use of multi-modal studies as a planning process to evaluate alternative transportation measures was reviewed principally drawing upon UK experience.

While transportation planning is well suited to the application of SEA, it is not without difficulties since the increasingly fragmented nature in which transportation infrastructure and services are provided in market-led economies results in a severely constrained command and control mechanisms within transportation planning institutions.

Tiering is one of the key concepts linked to SEA. This foresees that assessment is carried out at different levels of the decision-making level in a quasi-hierarchical way. The expectation is that in this way the scope of project EIAs can be better circumscribed and hence SEA can shorten project delivery timescales. The paper presented the view that this idealised approach is seldom likely to occur in practice, at least in the transportation sector and that many of the savings may not materialise in the manner intended.

Corridor Assessment and its Relationship to Projects
David Sweet, ISPA, European Commission

ISPA - the Structural Instrument for Pre-Accession Assistance - is a new instrument introduced by the European Union to help prepare the Central and Eastern European Countries that have applied to join the European Union. The funding available - just over one billion euro per year - is allocated to specific projects, each costing over five million euro, in the field of environment and transport. The major priority for transport projects is to support the Corridors identified in the TINA process.

This presentation reviewed the process by which projects are identified and developed, and discussed the particular problems associated with assessing and evaluating projects which are components in an overall scheme in addition to, and sometimes rather than, being justifiable on their own merits. This is a problem insofar as both the definition of relevant projects and their assessment remains a matter of national responsibility. Besides better establishing the operational meaning of the 'corridor', assessment tools need to establish criteria that go beyond cost-benefit analysis. Also needed is a better understanding of the trade-offs between economic growth and competitiveness on the one hand and regional cohesion on the other as both are explicit goals of the 'corridor' concept.

Old Myths and New Realities of Transport Corridor Assessment: Implications of EU Interventions in Eastern Europe
Deike Peters, Institute for Transportation and Development Policy

In the 1990s, the advent of the European Union and the break-up of the Soviet Union have brought profound new challenges to all levels of planning and policy making across the European continent. A new, increasingly border-free 'Europe of Regions' is emerging, constituting a new, yet unfinished supra-national geography. Especially in the context of planned eastward expansion, the European Union now faces enormous pressures to uphold its self-proclaimed goals of harmonious, balanced development, sustainable growth, economic and social cohesion, improved quality of life and solidarity between member states. One of the most hotly debated issues in this context is the role of transport infrastructures in fostering regional development and cohesion.

There is a new theoretical consensus that large-scale investments into highways and high-speed rail networks in Europe have often failed to bring the economic development benefits traditionally associated with them. In particular, scholars increasingly warn of the so-called 'two-way roads phenomenon', showing that improving a transport corridor between an economically powerful core region and a weaker, peripheral region can in fact have detrimental effects on the latter. It is therefore important to think beyond Aschauer's (1989) public capital hypothesis and reconceptualize the complex relationship between transport provision and economic growth. A milestone in this direction was the British government-sponsored SACTRA report (SACTRA 1999). Yet these and other recommendations of a similar type have been ignored in the definition and implementation of both the TEN corridors inside the EU and the recently completed Transport Sector Needs Assessment (TINA) plan for the Central European Accession countries. Neither of these were, for instance, subjected to a Strategic Environmental Assessment (SEA).

Considering the tremendously important role the European Investment Bank has recently taken on in financing TENs and TEN extensions in CEE, one of the most important changes in this respect would be an institutional restructuring of the European Investment Bank to ensure that environmental and social assessment capabilities within the EIB become more comparable to international development banks such as the World Bank.

Key Issues in Transport Evaluation

Spatial Economic Impacts of Transport Infrastructure Investment
Jan Oosterhaven, University of Groningen

Estimating the spatial economic impacts of transport infrastructure is an unsolved issue that plagues economic science for a long time. The basic problem lies in establishing the economic development that would have occurred without the investments in infrastructure, in combination with the a priori uncertain direction of the causal impact of new transport infrastructure on the economic development of the regions or nations affected. In fact, infrastructure reduces the cost of both imports and exports of goods and services, the net effect of which is not clear. Macro economic research and macro economic models only give indications of the demand effects and the supply effects of bundles of historical investments in infrastructure. The outcomes, however, have only limited value when taking decisions on concrete infrastructure projects, since macro quasi production functions with infra-structure as one of the arguments do not easily distinguish between the type of infrastructure (road, rail, air, water, terminals, harbours) or between the different location where the same investment may be made. Surveys among firms and consumers also have various measurement problems, but do have the advantage of providing ex ante micro information. Among the major

disadvantages are, strategic answers, sample selection, and the inability to capture indirect effects on non-using firms or consumers.

Economic potential models for interregional infrastructure and integrated land use/transportation (LUTI) models for urban agglomeration infrastructure provide the most reasonable empirical answer to the question of the economic impacts of transport infrastructure investments. Both, however show one major defect, namely their unsatisfactory foundation in economic theory. A promising alternative is provided by the new economic geography models that are evolving into more broadly based spatial computable general equilibrium models (S-CGE models). There is however a long way to go before these models produce empirical results that are as solid as the potential and actual LUTI type of models.

Empirical survey literature, on the other hand, has produced some qualitative outcomes that appear to be of a more general nature. First and foremost, both S-CGE models and potential models show the basic spatial forms of the investments in point infrastructure and line infrastructure. Second, there appears to be a general agreement of the fact that new infrastructure only produce minimal economic effects in countries with abundant infrastructure services. This holds with one major exception. When new infrastructure resolves strong capacity limits in either point infrastructure or in line infrastructure the local economic effects will be considerable, but mostly at the expenses of cities and regions further out.

The Economic Development Effects of Transport Investment
David Banister, University College London
Yossi Berechman, Tel Aviv University

The paper argued that where there is already a well-connected transport infrastructure network, further investment will not on its own result in economic growth. Transport infrastructure investment acts as a complement to other more important underlying conditions, which must be met if further economic development is to take place. Additional transport investment is not a necessary condition, but acts in a supporting role when other factors are at work. These factors are the necessary conditions that need to be met if economic development is to take place.

An evaluation framework that considers these factors and the context sensitivity of transport infrastructure investment should place transport projects within the network context and prioritise investment objectives – to include transport benefits, issues related to double counting of non transport benefits, and the functional linkages between primary transport benefits and potential economic development effects. A proposal for a twin approach to evaluation was presented, based on a conventional cost-benefit analysis and a complementary analysis that takes a much wider view of the investment proposal which

explicitly relates the contribution of a project to a) the transport network as a whole; b) the value added of the project in terms of its contribution to local employment, the potential for increases in productivity, and the environmental impacts; and c) the distributional impacts in terms of the spatial effects on the regional and local distribution of services and facilities, and the social impacts.

Conceptual Foundations of Cost-Benefit Analysis: a Minimalist Account
Robert Sugden, University of East Anglia

The paper attempts to de-couple CBA from two assumptions that underlie much of neo-classical economics, but that have been subject to much recent criticism. The first is the consistency assumption: that individuals have stable preferences, and that those preferences satisfy the strong principles of completeness and internal consistency that are postulated in conventional decision theory. The second is the revealed welfare assumption: that an individual's preferences provide a measure of her well-being, or of what is good for her.

The consistency assumption is empirical, and therefore open to test. In the last two decades, there have been many controlled tests of conventional decision theory, and the results have generally been discouraging. Various predictable patterns of behaviour, incompatible with the conventional theory, have been found to be characteristic of most human decision-makers. Some of these patterns cause serious problems for CBA. One of these being the phenomenon of preference reversal that suggests that the mental routines that people use when choosing between two options are different from those that they use when putting a money value on a single option. The problems caused for CBA by such deviations from conventional theory have become more obvious with the growth of stated preference and contingent valuation methods, which try to infer preferences from survey data; but it is becoming clear that the deviations are not artefacts of survey methods: it is the theory, not the surveys, that is at fault.

The empirical failure of the consistency assumption also causes problems for anyone who wants to justify the revealed welfare assumption. There are strong reasons for expecting defensible judgements about individual well-being to satisfy certain consistency conditions. As a matter of empirical fact, a person's preferences may be unstable and context-dependent; as a matter of normative principle, we may respect the choices that result from those preferences (in the sense of not wanting to overturn them); but it seems wrong to say that the nature of the person's well-being is unstable or context-dependent. If judgements about well-being are required to satisfy consistency conditions which are not in fact satisfied by preferences, the revealed welfare assumption cannot be sustained.

The revealed welfare assumption has also come under attack from a different direction, from philosophers and philosophical economists who have set out to

analyse the nature of well-being (or – for those who start from a different meta-ethical position – the nature of widely-held judgements about well-being). The work of these theorists casts doubt on the credibility or persuasiveness of the claim that a person's preferences, even if internally consistent, reliably reveal what is good for her.

Welfare Basis of Evaluation

Marco Ponti, TRT Transporti e Territorio Srl

In line with its standard economic use, welfare is composed by, a) net social surplus (willingness to pay/willingness to accept minus the social costs of the resources consumed); b) distribution effects, i.e. gains and losses of different social groups. Since a parallel evaluation of the distributive context of taxation would be necessary in order to fully evaluate the latter type of impacts – and 'fiscal' evaluation is of extreme difficulty – the distributive content of a project/policy is often limited to some evaluation of its *direct* impact.

Contrary to the multi-criteria approach, an updated 'public choice' approach tends to *limit* as far as possible 'discretionary' values given by the decision-makers, assuming that the latter are not 'benevolent princes', and quite often have short-term, egotistic 'hidden agendas' (re-election etc.). The only areas where discretionary values are inevitable are the distribution aspects, environmental standards (especially when set at international level) and landscape or architecture-related aesthetic values. Nevertheless, 'switch' economic values are *always* to be made explicit within these areas. This means that the opportunity cost of achieving one objective against the other has to be calculated in order to guarantee *consistency* within the decision-making process.

Possible improvements on the standard CBA practice would cover (a) the marginal opportunity cost of public funds – if a project/a policy increases the state deficit, this has a *measurable* and *negative* welfare impact; (b) the 'option value' of the social discount rate – recent American work on the subject demonstrates that investing reduces the possibility of alternative choices over a long period, and this has an opportunity cost not only for the private sector, but for the society as a whole; therefore the discount rate has to be adjusted to the flexibility of the technology employed; (c) market failures in the transport-using sectors – most typical failures take place where marginal costs are below prices, as it is where "scarcity" or monopolistic rents prevail over neo-classical assumption of perfect markets; as changes in transport have a side effect on such monopolies, a standard CBA evaluation *underestimates* welfare gains from lowering transport costs.

The External Costs of the Belgian Freight Traffic: A Network Analysis of their Internalisation
Michel Beuthe et al., GTM-FUCAM

The strong expansion of freight road transports throughout Europe is an important source of congestion and pollution, as well as a cause of many accidents. It is most likely that this problem will only grow worse as it is expected that freight traffic will go on increasing over the coming years. This is not a problem which could be solved by recourse to a simple and unique solution, but which will require the conjunction of many different remedies. In some places a partial solution could be found in the building of enlarged infrastructures. However, spatial as well as budgetary constraints severely limit that kind of solution. A better spatial distribution of human activities should be encouraged, and various regulatory devices could somewhat contain traffic expansion and its invasion in urban environment. Another partial solution would be the promotion of transportation modes which have lesser negative effects, i.e. rail and waterway, and their intermodal combination with road, in order to substitute these modes to the use of direct road transports. One way to achieve this substitution could be to set up a pricing policy that would include the external effects of each mode.

The paper presented some results for the last measure. These results were obtained from a detailed modelling of the Belgian multimodal freight transport network within the overall trans-European network.

Assigning a Monetary Value to Noise Reduction Benefits: an Example from the UK
Iain R. Lake et al., University of East Anglia

The draft EC directive on the Assessment and Management of Environmental Noise launches the idea of EU-wide noise maps which should form the basis of action plans and strategies to combat noise pollution. However, such strategies need to balance the costs of noise abatement with the likely benefits. If such goals are to be achieved, then assigning a money value to noise would enable the benefits of noise abatement to be placed alongside the costs of noise control in a cost benefit framework. This paper demonstrated how money values can be placed onto noise, through use of a hedonic pricing study, and illustrated the valuation of road traffic noise as a case study. Hedonic Pricing (HP) is an economic technique for estimating the monetary value of noise based upon the amounts that people pay for houses experiencing different noise levels.

In the past HP studies have been criticised due to the time and effort required collecting data on all these variables for a large number of properties. This research demonstrated how these problems can be circumvented using a Geographical Information System and large-scale digital data. These

innovations permit robust HP studies to be achieved in a fraction of the time and cost, although the latter is greatly influenced by the availability of appropriate data. These innovations may permit HP to be widely adopted across the EU as a method of obtaining noise values. However, of crucial importance is whether the results, or those of other HP studies, can be transferred across individual countries or even between countries. One of the main limitations of HP studies is that different housing markets may exist in different areas and within these there may be different values for noise.

Accessibility Analysis Concepts and their Application to Transport Policy, Programme and Project Evaluation
Derek Halden, Derek Halden Consultancy

The paper described how accessibility analysis techniques are being developed for practical application in transport evaluation and was based upon recently completed research in Scotland.

Whether or not accessibility measuring techniques are also practical tools within transport evaluation, depends partly upon the policy and administrative approach to transport. In the past, there has been a complex mix of weak linkages between transport supply and demand; some managed through various semi-regulated private sector structures and some managed wholly through a political process. The general assumption, although probably impractical, was that transport supply could be maintained through public funding at a quality level which roughly met the perceived demands of the population. The main focus of transport analysis was therefore on transport demand.

However, throughout the EC, the demands of the population have increased to a level that cannot be accommodated in physical terms or matched by public funding, and this has been a key factor in creating pressure for transport policy changes. Within the new transport policies, it is recognised that strong policy linkages are needed between transport supply and demand, and that these should be based upon wider economic, social and environmental objectives. The paper looked at how accessibility measures provide the links between transport supply and these wider policy areas, like transport and economic development, social exclusion and distribution/equity issues, involving stakeholders in the evaluation process, network effects, welfare basis of evaluation, and time, term and uncertainty.

Involving Stakeholders in the Evaluation of Transport Pricing Policies
Rosario Macario, TIS Portugal

During the last decade the Urban Public Transport (UPT) sector in Europe suffered strong movements of change with three main goals: increasing its productive efficiency; reducing the gap between the price paid by the users and

the real costs of providing the service; and reducing public expenditure in the sector by introducing new ways to involve private finance.

To solve this problem pricing and financing schemes have emerged all over Europe during the last years and evidence exists that the implementation process is responsible for a considerable number of unsuccessful cases. These implementation difficulties arise from the difficulty of identifying winners and losers and consequently devising effective schemes to transfer the gains of the first into compensation to the latter. However the responsibility for failure at the implementation stage is also related to the not well focused evaluation of the different stages of the policy cycle for transport pricing and financing in urban areas. For many years the implementation of pricing policies followed a logically structured cycle of allocative decisions. This approach assumes a unitary decision making, or considering a group acting as a unit, and ignores situations of conflict, which arise whenever social activities are at stake and the different interests of societal groups are confronted.

Public acceptability is a complex problems as it requires the joint consideration of a number of scientific areas, in particular economic aspects, social, technological, legal and even managerial. Acceptability is an interaction between political effectiveness, here understood as the capacity of accomplishing the proposed objectives, economic efficiency in production and consumption, equity and social fairness and, last but not least, feasibility of implementation.

Appraisal and Evaluation of Mass Transit
Geoff Gardner, Imperial College

The need for a mass transit system presents developing cities with a major problem. The sheer scale of the demand requires large amounts of money, organisation, or more likely both. An urban rail mass transit system of the type most commonly built in the last decade can easily cost more than one billion dollars. The research tackled the very important question of why aid-funding might be used to implement an expensive, possibly sub-optimal, solution. It investigated the concept of ‘suitability’ and reviewed the theory of human problem-solving to examine the motivations of decision-makers in a developing city. Most importantly, the research sounds a warning that the scale and complexity of the issue, together with the power of vested interests involved, mean that on some occasions promises of intangible benefits will be used to promote prestigious schemes in very low income countries where the need for investment in basic health and welfare is greater.

The paper included the outline definition of a methodology that could be used as part of an improved appraisal process. This is based upon multi-criteria analysis and gathers together a very comprehensive framework capable of investigating every conceivable benefit of a system, even those currently

outside of standard appraisal practice. This represents a significant repository of collective wisdom, obtained from reviewed and refined with experts in lending agencies, universities and consultants. It is unlikely that there are major benefits or disbenefits of a mass transit system that would be overlooked if every part of this framework was examined, discussed and analysed.

It is clear that decision making is a complex process. So complex, indeed, that most people prefer to simply do their best and hope for a satisfactory outcome. One of the conclusions of the research is that decision-making, especially in developing countries has too many factors that prevent what technicians may consider to be logic and reason. Everyone has a different way of looking at things, especially when they come from different cultures. It is necessary to accept this from the outset and look for ways of improving the process within existing boundaries. A decision-making framework will not necessarily 'solve' a problem, but it will certainly facilitate dialogue and discussion in a guided manner that may lead to more understanding between the disparate parties involved.

Evaluation in Practice: Problems Encountered

Norwegian Urban Road Tolling: What Role for Evaluation?
Odd I. Larsen, Institute of Transport Economics, Norway

Toll cordons are now operated in the three largest Norwegian cities. This paper summarised the main features of the schemes and the main impacts on traffic that so far have been ascertained. The objective of the toll cordons is to raise money for road building and the design of the schemes reflects this fact. As expected the cordons have reduced the volume of car traffic crossing the cordons in the period of operation, but they seem to have had only minor impacts on the choice of mode.

In Norway very little was done with respect to the evaluation of the road projects they were intended to finance, but in general it seems that most of the projects are profitable. The cost of money raised by the toll schemes seems to be of the same order of magnitude as the cost of public funds in general, but with respect to design there is scope for improvements, especially by differentiating the rates by time of day.

The Oresund Fixed Link: What Role for Evaluation?
Steen Leleur et al., Technical University of Denmark

The Øresund Fixed Link was opened July 2000 after a construction period of six years. The link is 16 km long and connects Malmø in Sweden with Copenhagen in Denmark. The Øresund Link is one of the fourteen projects adopted by the European Council at the Essen meeting December 1994. The project received a record high financial support from the EU Commission equal to 191 million

EURO or 6% of the total construction costs. It is both a local/regional connection between two large cities across a national border and at the same time a strategic link between Mainland Europe (via the Great Belt Link finished in 1998) and the Scandinavian Peninsula.

The paper reviewed the major steps in the planning of the project and discussed the role of evaluation in this process. Special attention was given to the conflicts in the process and its main players. The review of the planning process forms the basis for an exploration of the possibilities for formulating new methodology that will be better suited to deal with the rather complex set of socio-economic consequences that can be associated with a large TEN infrastructure project such as the Øresund Link. Special attention is given to the modelling of strategic mobility and the types of information that can be provided to the decision-makers about the mobility improvements due to the new link.

Paris-Brussels-Köln-Amsterdam-London High Speed Rail: European vs. National Level Evaluation
Rana Roy, Independent Consultant

This paper reported on the problem of evaluation in cross-border projects which was identified in the course of advisory work for the European Commission on the PBKAL project. It is a peculiarity of cross-border projects that correctly specified national evaluations of their respective national sections will tend to generate an incorrect estimate of their overall economic profitability and, therewith, an incorrect estimate of the extent of public support which they merit. This is because most national governments, in order to determine the extent to which their section merits public subsidy from the national taxpayer, will, quite properly, seek to limit their recognition of the section's economic benefits to the share accruing to their own residents. The consequence is that the set of correctly specified national evaluations will tend to exclude from view a large stream of economic benefits, corresponding to half the international consumer surplus, ceteris paribus.

The solution to the problem depends on the presence of a decision-maker with an interest in, and responsibility for, recognising and securing this stream of 'supra-national' benefits. Clearly, in the case of the TENs, the European Union has an interest in recognising and securing these supra-national benefits in the national sections of any given project and in establishing a true measure of the economic return on the project as a whole. In the case of the PBKAL, therefore, the solution was to produce for the Union a new evaluation which integrated the formerly excluded stream of benefits: we called it the community component. The new evaluation of the PBKAL established that the community component constituted 27% of the corrected economic return of the project as a whole. With it, the project – as a whole and in each of its national sections – passed the relevant hurdle rates; without it, it did not. In turn, this trigged the national and Union subsidy decisions that have indeed enabled the project to proceed.

It also trigged the injection of one billion euros to the Commission's TENs budget, the first step in a multi-billion euro 'top-up' intended to secure the supra-national benefits of all the cross-border projects.

Against the background of a successful outcome for the TENs budget but a continuing failure of the TENs projects to overcome the problem identified, this paper critically reviewed the original analysis and identified a lacuna therein. What is missing is the prescription of a determinate process of non-discretionary decision-making by which supra-national European evaluations are mandated and supra-national European benefits secured.

5.3 Improving evaluation practices in transport; towards a better integration of political and technical perspectives

5.3.1 The agenda

The third and last TRANS-TALK workshop comprised four roundtables and addressed the following issues:

The first roundtable considered challenges to evaluation for specific areas, like network organisation, infrastructure investment, intermodality or road safety, and against this background discussed the guidelines for improving evaluation practices in transport developed by the TRANS-TALK network.

The second roundtable dealt with the problematic science-policy interaction in evaluation. Whether one talks about appraisal or ex-ante evaluation or about ex-post evaluation, a central difficulty is the science-policy interaction or the interface between technical and political perspectives in evaluation.

The third roundtable considered research priorities for the future in view of contemporary challenges to evaluation and recent developments in transport policy.

Finally the fourth roundtable addressed the wider question of challenges to transport policies and in particular the Common Transport Policy for the next decade. This took place at the European Parliament.

5.3.2 Roundtable: Transport policies, evaluation practices and new challenges

This roundtable was opened by ***Liana Giorgi*** of the ICCR who presented the TRANS-TALK Guidelines for Improving Evaluation Practices in Transport.

These guidelines, written in the form of a handbook, review the main principles of evaluation from the perspective of transport and discuss the most common problems in their application. A good evaluation design needs to consider the following dimensions:

a) the function or objectives of the evaluation;
b) the scope of the evaluation in terms of both the policy context as well as with regard to geographical coverage;
c) the timing of the evaluation with reference to the decision-process but also with regard to the types of effects under study (whether short-term or long-term);
d) the methods to be used by the evaluation in terms of data collection, data analysis and formal assessment;
e) the disciplinary make-up of the evaluation team;
f) the expectations of the client as well as the way the latter proposes to make use or exploit the evaluation results
g) the communication of results.

The handbook identifies several problematic areas for transport evaluation for which an improvement of background knowledge is called. These are: (i) the meaning of European-added value in the context of subsidiarity and multi-level governance characterising the European polity; (ii) environmental valuation; (iii) the question of transparency and related to this of the democratisation of decision-making processes; (iv) the role and operationalisation of multi-modal planning.

The roundtable deliberated the TRANS-TALK guidelines through a consideration of specific policy areas or instruments within transport.

Roundtable participants were invited to:

- provide an overview of best (and worst) practices in transport evaluation in specific areas and/or a historically informed perspective on how evaluation in specific areas has developed over the last years;
- portray and illustrate the contemporary challenges for evaluation in specific areas;
- reflect on the applicability of the TRANS-TALK guidelines for evaluation in specific areas and make suggestions for their improvement.

Bill Tyson of the Greater Manchester Passenger Transport Executive (GMPTE) and the Transport Management Group discussed network organisation at the urban level taking as case study the development of public transport policies in the UK since 1980, in particular with reference to the Transport Acts of 1985 and 2000. The ex-ante evaluation leading to the Transport Act of 1985 was near the worst practice end of the spectrum which was all the more regrettable considering that the Transport Act of 1985

introduced the most radical reforms in public transport (especially with regard to bus services) since 1930. The decision process that characterised the Transport Act of 2000 was fundamentally different with a more extensive role for evaluation: the objectives were far more clearly formulated; greater use was made of traffic forecasts, environmental trends and scientific evidence more generally; there was more use of specific targets and outcomes; consultation with stakeholders and the public at large was explicitly promoted; and there was greater use of non-governmental experts.

Challenges for evaluation in the field of network organisation at the urban level are: (a) the evaluation of broad policies conceptualised at the national level but implemented at the local level; (b) the evaluation of long-term impacts – the longer the time period for the evaluation the more difficult it is to separate the effects of any particular policy from exogenous changes; (c) clarifying distribution effects recognising that broad categories (like bus passengers) represent in fact largely heterogeneous groups.

Karl-Josef Höhnscheid of the Bundesanstalt fuer Straßenwesen (BASt) reviewed practices in safety related transport evaluations. Here the main problem remains that of establishing a clear relationship between the causes and effects of accidents on the one hand and the effects of road safety measures on the other. Isolating and quantifying the influence of a measure is often the critical point. Road traffic accidents are the most important input for the evaluation of road safety measures. Determining the cost of road traffic accidents comprehensively and in monetary terms (to the extent possible) is the other main issue in safety related transport evaluations.

With the improvement of knowledge concerning the effects on safety of various measures and the economic valuation of safety, efficiency analyses are becoming useful decision-support tools in policy making. Open questions concern: (i) The treatment of humanitarian costs – these cover mainly damages paid for physical and psychological harm to the victims and their families, lower educational and professional opportunities and loss of independence – and (ii) the treatment of extra market costs. There is no harmonised European framework available for the evaluation of road safety measures and this is desirable. The TRANS-TALK guidelines provide a useful general framework that would have to be tailored to the specific needs of safety related transport evaluation.

Sandrine Vanel of the Network of European Transport Researchers (NETR) reflected on the trends and challenges faced in intermodal evaluation. In most countries intermodality and combined transport are late entrants in the transport policy agenda but currently of high priority. Countries can be grouped into three groups with regard to their policies on intermodality. There are first those countries that follow a system approach to transport planning and which place an emphasis on setting priorities in accordance with a multimodal master plan

which capitalises on the modal interfaces and intermodality. Here we find Sweden, the Netherlands, Italy and Germany. In other countries – Norway, Denmark and France – intermodality is targeted but as a separate sector, quasi as an independent 'mode'. Finally there are countries like the UK which are dominated by unimodal planning and which have yet to make the shift to multimodal planning and intermodality.

Intermodal evaluation should not be thought of as distinct from evaluation for other modes. However it does face specific problems. One is that of definition. There is still no widely accepted definition of intermodality. Related to this are problems of data collection and data analysis. For example, with regard to data collection, intermodal evaluation faces the problem of the weakness of the 'common denominator' as it has to rely on different sources for different modes. Furthermore the data are often not suitable for integration. To this should be added the growing problems for planners and evaluators posed by the 'privatisation' of (now) commercially confidential data – a particular problem in the intermodal area, although, of course, more general.

The biggest challenge for intermodal evaluation is perhaps that of operationalising the system logistics model to transport management across all modes.

Michael Gommers of ECORYS-NEI in the Netherlands reviewed practices in the evaluation of infrastructure. The most common forms of appraisal in use in EU Member States are cost-benefit analysis and multi-criteria analysis. Yet despite the fact that EU member states use more or less the same appraisal methodology, the impacts to be included in the method and the way they are treated differ between the different countries. Thus whilst there is a significant degree of agreement on the direct impacts to be included and the appropriateness of monetary valuation – the main criteria parameters that are common to all countries include time savings, accident reduction, construction costs and vehicle operating costs – the range of environmental impacts varies considerably across countries and the question of the latter's monetary valuation is far from clear cut. The coverage of socio-economic or indirect impacts is the area where there is least agreement, with some countries excluding such impacts and others including a wide range. The contribution contains a comprehensive categorisation showing the full range of possible impacts.

The main challenge for infrastructure evaluation in the future is how to capture the community or network effects (NE) in TEN project appraisals which for the most part are still carried out at national level. The identification and assessment of such effects is principally held back by the lack of adequate modelling tools and data that can provide detailed information. The same applies to the concept of European Value Added (EVA). In TENs EVA refers to the contribution to network improvements which result in increasing

interconnectivity and interdependence among regions thus providing opportunities for additional economic activities and social integration. Further research into EVA and NE could improve future evaluations of TEN projects since the existing research literature on these topics is still limited.

Another challenge lies in the comparative assessment of public private partnerships and public arrangements of new infrastructure investments. Finally with regard to the development of comprehensive multi-modal evaluation methodologies, the challenge here lies in how to treat the rather different characteristics of the different modes in a consistent manner.

Viviann Gunnarsson of the Swedish Agency for Administrative Development (SAFAD) gave some insights on evaluation practices from the Structural Funds. SAFAD was in 1999 commissioned to analyse evaluation and procurement processes and also the organisational structure and the division of responsibilities for the evaluation of structural funds. The overall aim of the analysis was to achieve better prerequisites for improved and more efficient evaluation of structural fund programmes prior to the next programme period. SAFAD's meta-evaluation showed that whilst the majority of evaluation studies of Structural Fund programmes or projects carried out by different Swedish agencies was satisfactory, there were also several consistent deficiencies. These concerned primarily the description of the evaluation's orientation and methods and the programme's links with other programmes and initiatives in Sweden and the EU. Few evaluations were considered to be fully acceptable or exemplary in all respects.

The discussion centred primarily around the TRANS-TALK guidelines and their applicability. A summary of this discussion as well as other comments submitted to the TRANS-TALK guidelines through the electronic open consultation procedure can be read at www.iccr-international.org/trans-talk/consultation

5.3.3 Roundtable: Towards a Better Integration of Technical and Political Perspectives

The second roundtable was opened with a presentation of ***Francis Terry*** of Social Research Associates and the London School of Economics who explored the relationship between apparently value-free techniques for evaluation of transport projects and the political process. He exemplified his arguments with three case studies: the Channel Tunnel Rail Link; the South Yorkshire Supertram and the UK highway construction programme. Francis Terry argued that it would be naive to deny the political interest in evaluation (especially ex-ante or appraisal) just as it would be wrong to refute the public interest in it. The political interest in evaluation is as such not problematic. What is problematic is when this is one-sided, part of vested interests or not

accountable. This is also why evaluation ought not to be the monopoly of governments. The increasing role of universities / research organisations, professional bodies, of the local tier of government as well as of voluntary organisations or pressure groups contributes to better evaluation insofar as it leads to a diversification of transport expertise and supports openness, debate and accountability.

The roundtable discussion focused on the theme of evaluation culture – the set of structural / organisational characteristics and institutional practices on both the demand and supply sides that are key for an effective evaluation.

Catharina Sikow-Magny of the Directorate General for Energy and Transport of the European Commission reiterated the observation that in transport, appraisals or ex-ante evaluations are more common than ex-post and more often than not relate to infrastructure investment. She observed that the new Communication of the Commission on the Common Transport Policy (CTP) will place a stronger emphasis on evaluation, including on monitoring and on ex-post evaluation. She however also noted that the diversification of expertise, which the Commission has always supported through the Framework Programmes, creates certain problems when it comes to using the knowledge provided for the purpose of drafting policy documents like the forthcoming White Paper on CTP. She gave as example the case of the forecasts for transport growth which range between 0,5 to 3 per cent. Especially for road, the various traffic models elaborated by the research community for the Commission are very contradictory.

More generally the elaboration of policy at European level would seem to be closely bound to negotiation and discussion, primarily with the Member States. This is a lengthy process which often goes against the day-to-day interests of policy makers or administrators at either the European or the national levels. However it is more likely to lead to success and full implementation. Seen from this perspective it might not be as inefficient as it might look at first sight. As an example she referred to the ongoing discussion on the directive on railway pricing.

Rainer von Leoprechting of the Evaluation Unit of the Directorate General for Budget presented the work of his unit on evaluation standards. Besides being involved in the evaluation of Community programmes, the Evaluation Unit of the European Commission is in charge of advancing recommendations to other operational Directorates General on how to set up and run internal evaluation procedures that can assist them in the formulation and implementation of policies. A typical evaluation function has the following characteristics: it anticipates the decision-making needs in the policy areas of its Directorate General; it establishes an evaluation plan; it timely contributes to decision-making; it supports quality of ongoing management; it maintains its balance between requests of beneficiaries, managers, the superior hierarchy and the

decision-makers, i.e. it assumes a moderating role. The commissioning or use of external expertise is an important input to policy formulation. However the latter is a deliberative and as such a participatory process and this must be recognised by the internal evaluation units of the operational Directorate Generals.

Kristian Colletis-Wahl of the French evaluation society in France gave his views about the flaws of the current evaluation system based on experiences made working as evaluator in France. He called for more realism concerning the outputs of an evaluation and more attention given to the terms of reference, evaluative questions, criteria and indicators on the demand side (i.e. commissioning agencies). In many cases evaluative questions are based on the belief that evaluation is an exact science, which it is not. On the other hand, evaluators are equally to blame about the flaws of the current evaluation system in transport. Their principal mistake is that they devote little time to the elaboration of the evaluation design and the evaluation methodology. This applies in particular to the evaluation of infrastructure investment which often assumes a deterministic relationship between infrastructure and development. The latter, however, is not substantiated by evidence. Non-deterministic approaches may be a step towards better practice in evaluation. Public managers and other stakeholders may not appreciate the prudent statements concerning the impossibility to evaluate net impacts, nor the argumentation about the necessity to consider circular causality rather than cause-effect relationships. However the adoption of a non-deterministic approach to transport evaluation, hence also the shift away from positivism, need not be equated with giving up on theory, evaluation or more generally normative science. The construction of a non-deterministic representation requires the shift from a representation of infrastructure as an exogenous resource, the presence or absence of which conditions development, to a representation of development as the result of a process in which infrastructure as well as other elements play a part.

Jens Kromann Kristensen of the Public Management Service of the OECD talked about the role of evaluation in outcome focused management. First however he reminded the roundtable that neither evaluation as a science nor the discourse on the usefulness of evaluation is new. Indeed, though evaluations still abound it is no longer *the* instrument in the public management toolbox for effectiveness. Results-based management and budgeting is, in fact, more widespread. Recent developments in OECD member countries hint that the two approaches – evaluation and results-based management – might be meeting with the introduction of an outcome focus in budgeting and management systems.

Traditionally, governments have build their budgeting and management systems focusing on costs and inputs. Budgets have been structured according to institutional boundaries and typically specified how appropriations had to be

allocated to different inputs - i.e. salaries, building, equipment management and management controls centred on avoiding fraud and tracking how money was spend. Since the late 1980s- to mid 1990s, however, the focus has increasingly moved to *outputs*. This reorientation has at its core that public sector producers must report on how they perform, that targets for their performance must be established and that they must be held accountable for their performance against these targets. The output approach has however a number of limitations, not least that neither the public nor politicians think in terms of outputs but rather in terms of outcomes (on welfare, society, the economy etc.) Output management is also not very supportive of institutional learning in terms of policy. It is for this reason that there is currently a shift to outcome-focused management.

Outcome-focused budgeting and management involves formulating, implementing and evaluating policies, taking into account the need to establish the linkages between costs, inputs, process, outputs and outcomes, thus also evaluating policies both with regard to efficiency and with regard to effectiveness. Experience shows that outcome-focused management and budgeting is not possible without deliberation as it requires bringing together organisations involved in policy formulation, policy executing and audit or evaluative institutions. In that it necessitates enhanced co-ordination.

Finally ***Pekka Rytilä*** of the Finnish Association for Transportation Planning talked about experiences with transport twinning between Finland and Poland. These brought very clearly home the context sensitivity of evaluation or related activities. Inter-sectoral co-ordination identified as important for outcome-focused evaluation, for instance, is very difficult to achieve in CEEC countries which are still heavily sectorally organised. Along the same lines, privatisation capabilities in transition countries have been overestimated, structural problems underestimated.

5.3.4 Roundtable: New challenges for the CTP and research needs

Two speakers opened the third roundtable on new challenges for the Common Transport Policy and research needs.

Rijk van Dam of the Committee on Regional Policy, Transport and Tourism of the European Parliament pointed out that effecting a shift towards more environmentally friendly modes of transport ought to be the highest priority of transport policies in the future. Combined transport should therefore be promoted both through policy and in research. Currently the average speed of rail (including waiting time) is 19 km / hour; inland shipping is not much faster. Short sea shipping better combined with inland shipping is one possible solution to the pressing congestion and environmental problems and deserves to be

investigated as such. With regard to sectoral organisation, rail enterprises ought to place more of an emphasis on organisational / management reforms and interoperability, their ultimate objective being to improve the performance of rail with regard to punctuality, which is more important than the length of journey. It is equally important to remove barriers to competition for rail but also for shipping – unlike with road, the markets for rail and shipping are not operating. Finally whilst progress was made with regard to meeting environmental targets, more needs to be done in this field.

John-Hugh Rees of the Directorate General Energy and Transport of the European Commission talked about the role of strategic research for the Common Transport Policy. The priority areas of the Common Transport Policy in the coming decade will be: re-balancing the transport system; maintaining the modal split at 1998 level; reducing congestion; reducing the contribution of transport to CO_2 emissions; and reducing the number of accidents. New research should contribute to the enlargement of the basic data and information necessary for monitoring policies and for forecasting trends, including on mobility trends and the role of demographic patterns. Also needed are policy assessment tools which are stronger on quantification, vehicle movements capacity assessment, as well as sectoral analyses on the subject of pricing.

Unlike with previous Commission Communications on the Common Transport Policy, the forthcoming White Paper will seek to establish quantified targets for policy. Thus whilst it is intended to maintain the growth of the transport system (to meet the forecasted increased demand), its structure / organisation should be changed in such a way as to avoid or reduce negative effects: this means effecting a reduction of the rate of growth of road transport (especially in urban conglomerations) but an increase of the rate of growth of rail transport (up to 40 per cent). Such a re-organisation is expected to restore the modal split of 1998 by the year 2010. In the case of air, despite the latter's negative environmental effects no decrease is forecasted or promoted because of the mobility / leisure trends and the role of air transport in this connection. Nevertheless an attempt will be made to change the operating patterns of airlines and to effect some shift from air to high-speed train.

Christian Chenez of the Ministry of Public Works, Transport and Housing in France talked about the French transport research programme PREDIT2 (1996-2000) and subsequently gave the view of his Ministry on strategic transport research in the Sixth Framework Programme.

The evaluation of the PREDIT2 programme (in anticipation of PREDIT3 to run from 2001 to 2005) has shown that research funded under the programme has covered the fields defined as priorities, exhibiting overall high scientific quality. Very importantly the programme has contributed to making strategic alliances between industry, SMEs, academic, research laboratories and local authorities. Recommendations for improvement concerned programme and research

management, the link to transport policy concerns and networking. With regard to the latter, France is looking for synergies between the national transport research programme and the European research programme as well as for an amplification of the networking effect to cover both the air and maritime modes (which tended to receive the less attention in PREDIT2).

Christian Chenez regretted the lack of visibility of transport in the propositions currently being made (by the European Commission) on the Sixth Framework Programme (FP6). Many Member States are advocating for more transport research in FP6 unilaterally as well as multilaterally. In fact, France has endorsed the ideas presented in a position paper prepared by DETR (UK) in December 2000 and has joined an ad-hoc group comprising next to UK and France several other European countries. Besides 'lobbying' the European Commission to include more transport research in FP6, sectoral ministries in various Member States are finding out that it is as important to lobby their own research ministries along the same lines.

Chris Fox of the Department of Environment, Transport and the Regions (DETR) of the UK expanded on the substantial recommendations of this multilateral initiative. He pointed out that there is a need for transport research at European level that looks into the operational, fiscal and legal aspects of transport as well as technology. Transport research or research on sustainable mobility ought not to be integrated into the programme on sustainable development but instead kept as a distinct programme of its own. Urban transport should be part of this programme and not part of another programme or key action (as in the Fifth Framework programme). Research priorities comprise:

a) The efficiency and effectiveness of transport policies, paying particular attention on the role of demand management instruments, pricing, the enforcement of charging, interoperability, as well as the best use of infrastructure and available capacities.

b) Environmental impacts – how to best achieve the Kyoto targets and ensure quality of life through technological solutions (road vehicle engine, aircraft engines) but also environmental management guidance.

c) Safety, giving emphasis on urban road safety (especially for pedestrians and cyclists), public transport safety as well as the human-machine interactions.

d) Transport, society and the economy – understanding the need for travel and transport (modal) choices across population groups and developing in this connection a social inclusion agenda.

e) Transport technologies and innovation – even though technological advances in transport are slow in the making (and in diffusion) it is important for research to look out for innovative solutions that might end up

resolving a significant part of the negative effects of transport in the medium- to long-term.

The importance of ensuring that transport research remains on the agenda in the Sixth Framework Programme – and ideally as a separate autonomous programme – was also underlined by the Swiss representative ***Barbara Schaer*** of the Federal Office for Spatial Development (DETEC) and by the representative of the European Council for Automotive R&D (EUCAR) ***Ingmar Ackermann***.

The Swiss representative pointed out that aviation and especially its environmental impacts should be given more attention in FP6. In this connection new technologies for other modes – explicit mention was made of high-speed rail and magnetic levitation – as well as alternative means of transport ought to be further explored. The dynamics of leisure transport, besides freight transport and urban transport, deserve likewise more attention. The international dimension of transport must also be explored further considering that national action often leads to distortions. There are also several research areas that have already received attention in previous research programmes but which deserve to remain high on the research agenda of the European Commission as they are far from being exhausted. This includes research into the development of indicators and methodologies, modelling, the theme of enlargement and transport accounting.

Ingmar Ackermann discussed the proposed instruments for funding in the Sixth Framework Programme which in many respect differ from the existing instruments in the Fifth Framework Programme. Three 'new' instruments have been proposed by the Commission: networks of excellence, integrated projects and the networking of national programmes (the so-called article 169 provisions).

With regard to the first two instruments – networks of excellence and integrated projects – the distinguishing characteristic here is that the networks or projects to be supported by FP6 are meant to be larger and longer, the idea being to thus enable the better clustering and integration of related activities. Notwithstanding that the latter is a legitimate demand and can be understood as an attempt by the Commission to increase the policy relevance of research, caution is also called for. Conceived of as complementary instruments (to the funding of smaller scale research that is completed in a shorter period of time) these proposals must be welcomed. As substitutes they would not make sense because there is still a need for smaller projects carried out fast. More specifically with regard to integrated projects, it should be kept in mind that these raise serious demands on co-ordination and management and also questions about intellectual property rights.

Insofar as article 169 is concerned, there is here a large need for clarification. Closer synergies between European and national research are meaningful, especially in areas where there are national differences. However this article should not lead to the disappearance of European research as an autonomous area. This is important both politically and for science policy reasons.

5.3.5 Roundtable: The common transport policy, european-added value and subsidiarity

What are or should be the priorities of the Common Transport Policy and what is the role of the European institutions? How democratic are the existing institutional arrangements in transport policy and how can they be improved? What are the prospects for evidence-based policy and monitoring? These were the questions addressed by the fourth roundtable of the TRANS-TALK workshop.

Alfonso Gonzalez Finat of the Directorate General Energy and Transport of the European Commission noted that increased competition in the transport sector has both benefits and disbenefits and these must be better understood. The principal challenge for the CTP in the new decade is effecting a charging / pricing system that is transparent and in that does fairness to different modes whilst going some way in dealing with the problems faced by the transport system in terms of congestion and environmental externalities.

Michael Walsh of the Department of Environment, Transport and the Regions of the UK identified four priority areas for CTP: liberalisation, pricing, sectoral integration and social/economic integration. With regard to liberalisation, he noted that the contemporary discussion focuses primarily on freight transport, more attention ought to be given to passenger transport. The experience made with liberalisation in the UK should be processed and used. State-aid remains an unresolved issue. On the subject of pricing, he noted that what is lacking is confidence in the values / margins being proposed for different modes and types of effects. The problem of subsidiarity is raised strongly thereby. Public acceptance likewise remains a problem. With reference to sectoral integration, the important issue here is environment with the obvious links to pricing. However the wider issues of demand management (especially in urban transport) deserve closer attention. The question of decoupling is key to the integration of social and economic concerns. However it remains largely unclear how best to achieve this and the evidence is weak. Insofar as the social exclusion / inclusion agenda is concerned, it is important to ask to what extent social exclusion is distinct to the transport agenda.

Gunnar Eriksson of the Ministry of Industry, Employment and Communications in Sweden took up the issue of the modal split and the target of re-balancing

this to the 1998 situation. He concluded that the usefulness and applicability of such targets can be questioned. It is important, he claimed, to keep the consumer in mind. The consumer perspective prescribes looking for policy instruments that increase rather than decrease the options to individuals, or choices. Typically, economic instruments are well suited for this purpose. Deregulation and market opening are as important.

Pierre Perrod of the National Committee for Transport in France pointed out that the real challenge in transport today is that of growth so that we may even talk about a crisis of growth. The forecasted transport demand is so explosive that it makes the qualitative upgrading of services or changes in the system extremely difficult. To that should be added four major obstacles, the importance of which is likely to increase in the future: the environmental constraints, constraints in terms of space, financial difficulties and the organisational barriers inherent in the variable sectoral organisation. A more radical long-term approach (over twenty rather than ten years) needs to be combined with shorter-term more pragmatic solutions that encourage voluntary actions from transport actors across modes in the right direction, and with a more critical approach to current policies – for instance, are we really convinced that liberalisation is the right answer?

Wolfgang Doerries of the Ministry of Transport, Building and Housing in Germany pointed out that part of the problem in developing new policies or a new direction in policies at the political level has to do with the failure to hint at assessment studies and their results. Characteristic is also that policies or programmes are introduced but then not followed consistently in the long-term. We should therefore not be surprised if they then do not turn out as successful as expected. This is for instance the case of the TEN extensions to the East, which appear no longer to be as high a priority as they were in the nineties. Another example is the directive on strategic environmental assessment. The Commission needs to play a greater role in policy formulation and implementation, yet flexibility should also be part of the system in the latter respect. How to best achieve this still remains unclear.

On the subject of the democratisation of decision-making in transport, speakers agreed that this is an important issue. ***Alfonso Gonzalez-Finat*** referred to this as a new culture of decision-making brought about by the consolidation of the Union's own institutional framework. A White Paper on Governance is currently under preparation by the Commission – this is expected to provide some answers to the challenges faced in this respect. ***Michael Walsh*** underlined that popular support is crucial for the implementation of policies but that the lack of this is often the result of the lack of openness, transparency and accountability in decision-making procedures. The involvement of the private sector in implementation, for instance in investment, raises new issues and new difficulties in this connection. It is also important to recognise that conflict is endemic in policy-making and that it would be naive to think it can be done

away with. ***Gunnar Eriksson*** and ***Pierre Perrod*** underlined this too. Rather it is important to establish mechanisms for representing conflicting interests and in that recognising their legitimacy. Following on this ***Wolfgang Doerries*** pointed out that conflict, and, in turn, participation, needs to not do away with the demand on government to make choices.

With regard to the question of monitoring and evidence-based policy, ***Alfonso Gonzalez Finat*** noted that there are two aspects to this: first there is a need to improve monitoring and assessment tools; second, the results must be used. In the UK, noted ***Michael Walsh***, there is currently much ex-post evaluation going on – it remains to be seen how it is used, but also whether it is published, thus contributing to the afore-mentioned problem of transparency and accountability. When discussing the harmonisation of tools or assessments, one should nevertheless be aware that this could at the same time imply a levelling off which would not be consistent with democracy. ***Gunnar Eriksson*** called for more attention to be given to benchmarking exercises, but also for caution with regard to the development of common harmonised monitoring systems. Instead attention should be directed towards defining a minimum set of inputs that are standardised and common to all. Monitoring mechanisms, according to ***Pierre Perrod***, must consider the decision framework, i.e. who decides and what, who finances, who regulates and how integration / enlargement is to be achieved.

6 Annex B. Guidelines for Improving Evaluation Practices in Transport

Improving Evaluation Practices in Transport

Guidelines prepared in the framework of the

TRANS-TALK

Thematic Network

Project and Policy Evaluation Methodologies in Transport

supported by the

Transport RTD Key Action ‘Sustainable Mobility and Intermodality

of

The Fifth Framework Programme

Preface

The TRANS-TALK network was established as a forum to deliberate on the problems and prospects of transport evaluation in the European context. Three seminars were organised in 2000/2001. The first dealt with the context, theory and methods of evaluation, the second with the needs and capabilities of traditional evaluation methods when applied in the transport sector. The third seminar discussed a preliminary version of these guidelines.

The purpose of this handbook is to provide guidelines that can help better design evaluations of policies, programmes or projects in transport. These guidelines are not meant to substitute existing evaluation manuals at national or European level, but rather to extend them and help improve their use. It is also not the handbook's intention to set European-wide technical standards with regard to evaluation methodologies in transport – for instance in relation to the valuation of impacts in cost-benefit analysis, the weights applied to criteria in multi-criteria analysis, or the basis of projections and forecasts for transport demand and supply scenarios. Rather this handbook provides a framework that assists in the development of such standards, if the latter are judged necessary or desirable.

This Manual will be useful for evaluators working in the transport sector – both those working for public policy institutions and those external to the latter, yet commissioned to carry out independent evaluations.

Definition and use of evaluation

Evaluation, many manuals recognise, is defined (slightly) differently by different scholars. This in part reflects the more general policy context and purpose which gives rise to its need. In other words, how evaluation is defined stands in close connection with its use. In all definitions evaluation is about the assessment of a certain (policy) programme or project (or a set of programmes or projects) whereby definitions differ with regard to the main reference of this assessment, including for whom or for what purpose the evaluation is considered good for.

Examples of the main uses and definitions of transport evaluation are:

- Evaluation defined with reference to high-level objectives like economic efficiency is usually commissioned to look at the long-term outcomes of a policy, programme or project from the welfare maximisation perspective.

- Objective-led evaluation commissioned by the management level of an organisation or policy institution is more concerned with exploring the use of funds with reference to the priorities set by the organisation.

- Evaluation defined with reference to the impacts of a policy, programme or project (overall or a specific sub-set) is concerned with identifying the short- and/or long-term effects across several dimensions.

The use of evaluation is also influenced by who is commissioned to carry out an evaluation. Internal evaluations are often carried out as self-evaluations to support learning and improve implementation. Evaluations by external experts, on the other hand, are more appropriate when the objective is to provide new perspectives on public policies or when there is a need for specialised evaluation skills. Finally, independent evaluations might be more appropriate when the objective is to increase transparency and accountability.

The analytical framework

The analytical framework of evaluation describes the set of premises and assumptions entailed in the evaluation design with regard to:

1. The definition of the issues or problems under investigation, their interrelations, their degree of complexity and institutional setting.

2. The definition of (social) welfare and, in this connection, the weights assigned to issues like efficiency, economy and effectiveness or considerations about equity and ethics.

3. The choice of data and methodologies.

The explicit reporting of the premises and assumptions with which an evaluation design operates is important for judging the relevance and timeliness of the evaluation; and for deciding on the scope of the evaluation and the methodologies to be used.

An example can be provided by considering the question of accessibility – say the evaluation of the effects on accessibility of the transport-specific projects of the Cohesion Funds. It is first important to define accessibility. Whether accessibility is defined in terms of geographical peripherality (measured in terms of distance or travelling time to central regions or towns) or in terms of structural economic development has significant implications for the evaluation design. The approach to be taken, and likewise the indicators to be selected, will also differ depending on whether the emphasis is alone on economy and efficiency and/or equity or, more specifically, the gap between richer and poorer regions.

Scope of evaluation

In evaluation targeting public policy expenditure, a distinction is often drawn between policy, programme and project evaluation. This distinction is analytical and useful for describing the degree of complexity of the subject matter.

The term ‘project’ is usually used to describe an initiative which is strictly delineated in terms of time, space and budgetary allocation. A ‘programme’ refers to a set of interrelated initiatives which extent over time and space, in parallel or in phased manner with changes possible at various stages. Policy evaluation as a term is then retained for describing either policy analysis as a multi-faceted discipline or analysis for policy, i.e. analysis that helps formulate policy consistently.

Some evaluation manuals take a step further and argue that methods or tools of evaluation are differentially suitable depending on whether the subject of study is a project, a programme or a policy. Briefly quantitative or modelling techniques are thought more suitable for project evaluation, qualitative techniques more suitable for programme and/or policy evaluation. There is, however, no set standard on this issue.

Defining the scope of evaluation using the P/P/P (policy, programme, project) terminology is useful for describing the complexity of the subject matter, yet whether the latter is a policy, programme or project cannot be determined independently of the use to be made of evaluation.

For example: An infrastructure investment plan to link A to B through rail might be a strictly delineated project, yet it at the same time is part of a wider rail network and in turn of a multi-modal network linking the two regions where A and B are to be found. Depending on the objective of the evaluation, this wider context might play a greater or lesser role, thus shifting the boundaries of the subject under study, hence also of the scope of evaluation.

As important for determining the scope of transport evaluation is the area of intervention and the geographical scope of the P/P/P in question.

The following instruments are available to policy-makers at European level (in the context of the Common Transport Policy) and the national level:

1. Infrastructure investment
   Transport infrastructure planning, investment and financing;
2. Network organisation
   Bus lanes, speed limits, measures supporting intermodality and interoperability;
3. Traffic system management
   ITS, GNSS, fleet management systems, etc;
4. Standard setting
   Environmental standards, technical standards, etc;
5. Market regulatory framework
   Market access, deregulation, liberalisation, etc;
6. Pricing and taxation
   Road pricing, fuel taxes, charges for parking, etc.
7. Human capital / factors
   Education and training, Research, technology and development.

With regard to geographical scope, it is important to distinguish between:

- At the national level:
  Urban level
  Inter-urban level
  Regional rural
  Inter-regional rural
- At the European level:
  Regional cross-border

Inter-urban cross-border
Trans-national cross-border
Trans-national network

The higher the level of abstraction be it in terms of type of intervention or in terms of geographical scope, the more complex the evaluation is likely to be. The most complex evaluation designs are those that address in parallel or jointly several intervention measures, and at different levels of spatial disaggregation – even prior to determining the objectives against which the project is evaluated.

Used alone, quantitative or modelling techniques are ill-suited for dealing with complex evaluation designs. Not because it is a priori impossible to model or quantify effects but, rather, because models are ultimately tools of simplification, resting on a (usually limited) set of assumptions or hypotheses. Often such simplification is useful and necessary. The point made here, however, is that there are areas of intervention that cannot be assessed with modelling techniques *alone,* without running high the risk of unreliability of results. In any case, it is important when planning to use models as one methodological input in complex evaluation designs to become acquainted with the assumptions on which these are based so that the results derived from their use are correctly interpreted.

Timing of evaluation

The time dimension is relevant for evaluation in two ways.

First, the timing of the evaluation as such with regard to the phase of implementation of the project in question is important. Whether an evaluation is undertaken at the planning or design phase, during implementation, or after the project has been completed fundamentally determines its function. Evaluation carried out during the planning phase – often referred to as *ex-ante evaluation* or *appraisal* – has the primary function of structuring existing information to deliver insights into the expected outputs, results or outcomes of the project. Evaluation carried out during implementation – often referred to as *mid-term evaluation* or *monitoring* – has the function of observing developments to deliver a preliminary assessment of the project's effects and/or of the extent to which it is delivering according to plan. Finally evaluation carried out once implementation has been completed – often known by the name of *ex-post evaluation* – is meant to furnish policy-makers with information about the results and outcomes of the project.

In the transport sector initiatives are known to take a long time to implement and phasing is a common occurrence. This tends to blur the above distinction. The

fact that comprehensive evaluations need themselves time to be carried out further complicates the situation: it is not uncommon to find an appraisal commissioned at the planning stage, with the objective of determining whether the initiative as planned is sensible, only being completed after the project has been decided upon. It is situations like this that lend support to those who argue against evaluations or who question their relevance.

The second way in which the time dimension is relevant for evaluation is the time horizon for which effects are to be observed or forecasted. This is related but not dependent on the time frame of the project under study.

There are three types of effects that can be the subject of an evaluation:

- *Outputs* describe the product of an activity or what is obtained in exchange for (public) expenditure. Outputs are by definition tangible and short-term.

- *Results* describe the short-term or immediate effects of an activity with reference to the immediate addressees, recipients or impact groups and the original objectives of the activity.

- *Outcomes* or *impacts* describe the mid-term or longer-term effects of an activity with reference to impact groups and generally valid objectives.

Depending on what type of effects stand under examination, indicators need to be defined. Result and impact indicators are often the same, however it is important to differentiate the time frame and in this connection take into account the time series or statistical distribution of impacts, as known from previous research. Employment effects of transport initiatives is a well-known case in point. The immediate short-term results on employment of a transport initiative can well be measured by the number of jobs directly created by the project in question. This same indicator is however of little use – on its own – for measuring the indirect or long-term employment impacts of a transport initiative. The latter can only be measured by proxy variables tapping on economic activity which can be shown to stand in relation with the transport project in question. Another example is that of safety impacts associated with the introduction of speed limitations. Yearly observation data on the number of accidents can be misleading unless corrected for statistical distribution effects.

In cost-benefit analysis, more specifically, the valuation of long-term impacts must also consider the question of discounting, which entails making conceptual choices about inter-generational equity, as well as the question of length of investment and that of possible irreversibility of effects (in particular relevant for the valuation of environmental impacts). In all of these cases it is important also to recall that any estimations made are likely to rely on input data that are uncertain, at least to some degree, hence sensitivity analysis is important.

Another problem with the study of outcomes is that the longer the time perspective becomes, the more difficult it is to separate real effects (related to the operation of a P/P/P) from exogenous effects.

In ex-ante evaluation or appraisal it is often not possible to do this without first specifying scenarios for the future. Such scenarios should include forecasts on relevant context or background indicators as well as an outline on how these context indicators relate to each other and to the impact indicators chosen. If it is assumed that the context indicators relate to each other and to the impact indicators in the same way as they do at present, then we talk about a 'business-as-usual' scenario. Often the 'business-as-usual scenario' is not enough as it does not do sufficient justice to ongoing or potential socio-economic changes. It is therefore important to have next to a 'business-as-usual' scenario at least one 'business-not-as-usual' scenario, preferably, a set of these. This also allows the carrying out of a degree of sensitivity analysis and represents at least some attempt to understand the likely impacts of uncertainty or risk.

Evaluation methods

At various instances in the previous sections reference was made to the fact that the evaluation design influences the applicability of specific methods and should thus influence the choices made.

Methods for data collection

- **Surveys**
- **Use of secondary data**
- **Existing information / databases**
- **Case studies**
- **Focus groups**
- **Natural observations**
- **Expert opinions**
- **Programme documents**
- **Literature reviews**

Methods for data analysis

- **Statistical analysis**
- ***Models***
  - **Input/ Output**
  - **Micro-economic**
  - **Macro-economic**
  - **Statistical**

- ***Non-statistical analysis***
  - **Expert panels**
  - **SWOT analysis**
  - **Colour vote**
  - **Benchmarking**
  - **Logical framework**
  - **Delphi survey**
  - **Group interviews**
  - **Meta-analysis**
- ***Formal assessment techniques / aggregation***
  - **Cost-benefit analysis**
  - **Cost-effectiveness analysis**
  - **Multi-criteria analysis**
  - **Scenarios**
  - **Impact assessment**
  - **Policy analysis**

Source: MEANS, 1999

Evaluation results are only relevant if timely delivered. This is also why in an ideal world it ought not to be the task of the evaluation team to collect anew data. *Surveys* as a method of data collection are often the best means for obtaining quantitative data from a large number of respondents and are useful for making comparisons, e.g. across time, settings or types of respondents. However, surveys also display potential disadvantages which need to be carefully considered. In order to ensure representativity and validity, extra care is needed when drawing the sample and when following up respondents to obtain a high response rate. The questions must also be thought through carefully to avoid systematic response biases.

Case studies are often used in transport evaluation in combination with secondary data. The problem that case studies present and which the evaluation team needs to be aware of is the degree to which the results are transferable. Therefore a careful selection of case studies is necessary.

Expert interviews are likewise often used to collect information in evaluation. As with case studies, care must be taken to select a representative if not an exhaustive sample of expert actors. Expert interviews can be used to collect information on available data; but most of the time they are used to survey expert opinion. As such they are most useful for evaluation designs that seek assistance with the interpretation of available information. *Focus groups* are likewise most useful for evaluation designs that examine decision-taking or consensus-building processes as they are a tool that can help reveal how actors are likely to perceive, represent and deliberate on specific issues.

The review of *programme documents* is used in evaluation studies as a source of information on the decision-making context and the client needs. Programme documents present a specific view on the project which is unlikely to be comprehensive. When programme documents are used, it is important for evaluation teams to seek and obtain official and unofficial material. They should in no way be the sole source of data or information collection.

Finally *literature reviews* in combination with the use of *existing databases* is probably the most efficient way to collect comprehensive information on the subject of study. Given however the difference in scope and objectives of previous research or existing databases, their use must be guided by a good conceptual framework. Specifically with regard to the use of existing databases, a key to their selection is their documentation. Undocumented or poorly documented databases are of little use as they provide no sound basis for data analysis and interpretation.

Whether quantitative or qualitative methods are used to collect information will depend on the scope and design of the evaluation and the subject of study. A methodology mix is, however, in any way advisable. Overall, it is important when selecting a method (or methods) for data collection to reflect both on the reliability of the method(s) and the validity (internal and external) of the results produced.

The cautionary remarks made with regard to the use of existing databases apply also to the use of *models*. Within any specific evaluation study, it is unlikely that there will be enough time to develop a model. Thus existing models will have to be used. Existing models that are undocumented or poorly documented represent more a problem than a solution and should be avoided. Documentation is also important because without it there is no basis to establish the underlying assumptions that are necessary for its use. It is the task of the evaluation team to validate the model's assumptions against those prescribed by the evaluation design for the project under study. Failure to do this can lead to unreliable results and their lack of acceptability.

The applicability of *statistical analysis* is dependent on the quality of the data available. If the evaluation design foresees the carrying out of an own survey to collect data, then it is important to seek statistical advice already at the stage of planning the survey. Many surveys have turned out to be of limited use (in relation to the input) because statistical advice was not sought early enough.

Non-statistical methods for data analysis are often thought to be most useful when the data to be analysed is of a qualitative character. Non-statistical methods are however also useful if used in conjunction with statistical methods or models as a way to structure the presentation of results; assist their interpretation; or receive feedback. As such they are indispensable tools for

dissemination and exploitation as well as for improving the representation of stakeholders' interests in the evaluation process.

The use of models and/or statistics to guide data analysis in evaluation must be informed by a good level of theoretical knowledge and relevant documentation. Assumptions made in the evaluation design regarding the subject of study must be consistent with those of models or statistical theory. Non-statistical methods can be used to complement statistical methods. Their merit lies primarily in their ability to capitalise on users' and stakeholders' interests and in their ability to communicate evaluation results in a policy-relevant and comprehensive way.

Formal assessment techniques are used to combine the results deriving from data analysis to arrive at overall conclusions in accordance with the objectives set for the evaluation.

Cost-benefit analysis is used when the objective of evaluation is to compare the costs and benefits of a project using a common denominator (usually money) in order to decide on whether costs outweigh benefits or vice-versa. The use of cost-benefit analysis entails acceptance of a series of assumptions regarding, for instance, social welfare (that this is the sum of individual welfare), individual welfare (that this is reflected in preferences and that in turn these can be measure by the willingness-to-pay or the willingness-to-accept), the role of the market (that it reflects WTP or WTA) and the substitutability of resources as well as of costs and benefits. Problems in the acceptability of cost-benefit-analysis results derive from fundamental disagreements within society as well as among policy-makers about the validity of these assumptions. Problems with the applicability of cost-benefit-analysis derive from measurement hurdles or disagreement about monetary valuation procedures.

Multi-criteria analysis is often presented as an alternative to cost-benefit analysis in cases where the majority or an important set of relevant effects cannot be monetised. An important, albeit not unproblematic, assumption entailed in multi-criteria analysis is that weights applied to different effects – to reflect their degree of importance on the impact scale – can be established with reference to decision-makers' opinions and that the latter also truly represent the views of consumers and citizens.

Both cost-benefit and multi-criteria analysis rely on *impact assessment* techniques as first input. Impact assessment techniques delineate functions that relate the initiative under study to result or output indicators. It is possible for an evaluation study to present merely the results of these transformations without seeking to aggregate these results through a comprehensive framework like cost-benefit or multi-criteria analysis. This is suitable especially for those types of impacts for which there exist thresholds to monitor policy performance or implementation.

The use and relevance of *scenarios* was addressed in the previous section. To reiterate, scenarios are indispensable when studying the long-term effects of an initiative. Scenarios describe visions of the future that make explicit how context indicators relate to each other and to impact indicators.

In policy science, *policy analysis* is a term used to cover all set of methods or approaches that can be used to make any form of judgement on public policy. As such it includes the work of economists – thus also cost-benefit analysis – as it does that of political scientists – studying decision processes – and sociologists looking at organisational aspects. Here policy analysis is used to delineate meta-analysis that seeks to integrate the results from the application of different tools to arrive at recommendations.

The choice of a formal assessment technique (or a mix of these) is one of the most difficult choices involved in evaluation. Unfortunately it is the one which is most of the times arrived at quickly. This is in part due to the professional bias of the evaluation team. It is however also not unrelated to the demands imposed by clients. In transport, cost-benefit analysis is still often considered the only legitimate way to arrive at a decision on a transport initiative (especially with regard to infrastructure investment). This said, the diffuse recognition of the limitations of cost-benefit analysis has over the years led to the adoption of multi-perspective evaluation designs where cost-benefit analysis is one rather than the sole form of input.

The assessment techniques to be used in evaluation must be chosen at the onset of the evaluation, thus influencing the latter's design. They should in principle inform data collection strategies or methods of analyses. A combination of impact assessment techniques with multi-criteria analysis or cost-benefit analysis is the typical approach taken in transport. Scenarios are indispensable when long-term effects are taken under the loop. Policy analysis as a form of meta-analysis can assist in the integration of results and the working out of strategic impacts.

Strategic assessment is a term currently in use to describe a decision framework that seeks to establish the context indicators necessary as input in the evaluation of specific initiatives. It is being proposed as an priori to environmental impact assessment but also transport planning and includes a set of expert assessments that describe the likely outcomes (on the environment) of different sets of policies. These assessments must then be taken aboard by project specific evaluation studies.

The proposed strategic environmental assessment directive in its current format is primarily expected to contribute to evaluation by making it mandatory to take a more global view to assessment, thus ensuring the integration of environmental concerns – and of various stakeholder interests – in evaluation. It remains an open question whether it may be useful without the establishment of

common context indicators (and subsequently thresholds) for measuring environmental damage.

The role of research

Where are the boundaries between research and evaluation to be drawn? Most evaluation manuals are keen to underline that evaluation is not research, evaluation being much more policy or project driven than research, the aim of which is to improve the socio-economic knowledge basis. This distinction, like many others, is not as clear-cut however.

Complex evaluation designs are often not possible to implement unless granted enough time and resources and unless they advance the state-of-the-art, be it by collecting new data or information, by applying existing methods to new fields of application, or by developing new assessment methods. Research on the other hand often comes up with results which are relevant for the evaluation of policies.

Perhaps the one remaining distinct characteristic of evaluation as compared to research is that the former unlike the latter is explicitly commissioned to provide input to policy deliberations, and often to specific policy clients. In that evaluation has a more direct connection to politics. Research on the other hand can claim more autonomy, but it often goes unnoticed by policy actors. No doubt, institutional specialisation – with universities or research organisations 'specialising' in research and consultants 'specialising' in evaluation – has contributed to the drawing of clear borders between research and evaluation.

However things are changing. A detailed exploration of the reasons behind this change are beyond the scope of this guide. Suffice to note that among the main reasons for this change are the change in funding structures of research institutions – including universities – away from core funding towards contract research and the recognition by the industry that basic research is as important or necessary as applied research. The blurring of the borders between evaluation and research is an opportunity for the better exploitation of research results, and can be used to build new institutional alliances between research and industry.

In any case, in modern evaluation it is necessary that the evaluation team is up-to-date with the state-of-the-art in the field under study. This will also allow it to determine the extent to which new research will be necessary for carrying out the evaluation or whether it is possible to rely on existing information. If new research is necessary, this must be taken into account in the time plan and budget.

Being up-to-date with the state-of-the-art in the field under study is however also important because evaluation is often seen by commissioning agencies as additionally a means to deal with information flow. Evaluation is often expected to filter information through for what is important for a policy decision. This can only be done on the basis of a good knowledge of the state-of-the-art.

The evaluation team

The composition of the evaluation team is very important for the success of evaluation. Besides expertise, evaluation manuals point to independence as an important criterion. Whereas this applies primarily to external evaluations, it is also a more generally valid point.

Evaluators should not have any direct or indirect personal or institutional interest in the subject of study. In relevant calls this is usually implemented by adding to the eligibility criteria for participation the requirement of supplying proof that the applicant has not in the recent past been involved in any activities relating to the project under evaluation. Failure to ensure independence can endanger the legitimacy and acceptance of the findings later on.

In modern evaluation the multi-disciplinary composition of the evaluation team is as important a criterion as proof of expertise and independence. As we saw earlier, evaluations often go wrong because of the failure to recognise possible pitfalls in the evaluation design. The more complex the evaluation design, the more difficult it is for one evaluator or a mono-disciplinary evaluation team to ensure quality control. Besides, a multi-disciplinary team guards against bias in evaluation which might arise even when no direct or indirect interest or stake in the subject of evaluation can be established.

The role of the commissioning agency

Gaining support from the top, generating effective demand, setting realistic expectations, systematising evaluation activities, and linking with budget process, are some of the key issues in improving evaluation practices.

Not all organisations or all policy areas are ‘evaluation friendly’. Indeed for some evaluation is a new challenge, often imposed by regulations – quasi from above – or recommended by external advisers as a means to re-engineer or modernise decision processes. Learning how to use evaluation is a process largely of organisational change.

Staff members at different levels need to be convinced about the advantages of evaluation and this might involve providing them with training – about what evaluation entails and about its added-value and implications for their work. With regard to public expenditure at sectoral level the decision to introduce evaluation often entails changing the way in which the policy-owner thinks about and communicates policies.

How evaluation agencies think about evaluation is an important facilitator or barrier to the successful implementation of an evaluation and the acceptability of its results. In commissioning agencies where there are no standardised procedures for planning and *processing* evaluations, evaluation results are of little use.

Involving stakeholders in evaluation and the role of the public

In modern evaluation and especially that commissioned to look into programmes involving public expenditure, the calls for the greater involvement of stakeholders and the better communication of evaluation results to the general public are getting louder.

Stakeholders represent individual or institutional actors that have a stake or interest in a decision taken at the policy, programme or project level, and who seek to influence this decision in such a way as to maximise their benefits and minimise their losses.

Involving stakeholders in the evaluation process does not mean manipulating the results in such a way as to benefit one or the other stakeholder, nor is it a means to achieve a win-win solution for all involved. This is indeed hardly possible in situations where there are conflicting interests, which is the situation in most decisions involving transport policies, programmes or projects.

Involving stakeholders in the evaluation process rather means:

- at the minimum level, rendering the conceptual and methodological choices which are inherent in any evaluation design transparent;

- at the maximum level, constructing the evaluation as a *process* for finding a solution which maximises the benefits for most whilst minimising the losses for the few.

The above problematic is best illustrated by an example. The assessment of the environmental externalities of road vs. rail towards determining fair prices is often complicated by the confrontation between stakeholders of either side – the road lobby argues that the road user is already paying for the externalities, the

rail lobby argues the opposite and talks about unfair competition conditions. A closer look at studies commissioned by either side reveal that the values attached to environmental burdens differ quite significantly, whereby there are often good reasons for either position. The results are also influenced by a range of other methodological choices, for instance with regard to the role granted to population density in the models used, the geographical scope (in terms of distance) or the types of environmental effects considered.

Making the conceptual and methodological choices transparent does not resolve the above conflict but it does provide a basis for reflection. This can then be the starting point for a sensitivity analysis to test the robustness of the results under varying conditions. Subsequently such analysis can be used to decide jointly on that option which comes closest to a Pareto-like optimum. Needless to say, such a compromise is only possible if the stakeholders involved recognise that there can be no win-win solution such that the benefits of all or maximised and there are no losses involved.

In other words, it is possible to involve stakeholders in evaluation, however how this is done and with what objective ought to be decided at the onset, and accordingly the evaluation design should be determined.

If the evaluation is designed to contribute actively to the deliberation process between stakeholders, then qualitative methods of data collection and data analysis like the dephi survey, group interviews, the colour vote, expert panels or focus groups will have to be used to accompany the use of quantitative or formal assessment techniques, like multi-criteria analysis, which may reflect stakeholders' opinions but which are not suited for deliberation.

More generally there is an increasing need to establish open and trusted communication between governments and the public. Transparency is here very important. Commissioning external and/or independent evaluations or evaluations that aim to contribute to deliberation directly are ways to overcome this mistrust. Promoting forms of direct citizen participation – through focus groups, planning cells or public inquires following similar rules to those described above for stakeholder involvement – are additional means to reduce public mistrust.

Evaluation design handbook and final report

Several evaluation manuals underline that for evaluation results to be useful they have to be reported well in the final report. Final reports must be comprehensive and take into account that the audience is a general one and in that not necessarily fluent in evaluation terminology or scientific jargon. Writing

a report for a general audience is for professional evaluators often more difficult than writing for their own peers.

Equally important however is that the evaluation team writes down equally comprehensibly their evaluation design at the outset of the evaluation. Besides helping to clarify open questions with clients, the evaluation design handbook provides a reference manual for the evaluation team throughout the evaluation process. This is in particular important when evaluation teams are composite and include members from different disciplines.

An evaluation design handbook may – and indeed should – follow general guidelines as proposed in generic evaluation handbooks, however it must also be tailored to the specific needs of the project under study. Thus whereas general evaluation handbooks remain necessarily broad-spectrum on the specific indicators and/or the measurement rods of formal assessment methods and give no recommendations as to the assumptions to be used with regard to the context and subject of study, the evaluation design handbook must be explicit on all these aspects.

The methodological choices and relevant assumptions must then be reiterated in the final report (preferably in the annex) with clear indications as to where these might have changed in the course of the study.

Golden rules for good evaluation practice

The following are golden rules for evaluation practice:

1. Specify from the outset whether the evaluation is understood as an assessment relating to generally valid objectives; an assessment relating to the specific original objectives of the programme or project under study; an assessment aiming to study the project's effects more generally; or a mixture of the above.

2. The scope of evaluation cannot be determined independently from the objective of the evaluation exercise. Nor can it be determined without due consideration of the background information available on the subject and the context of study.

3. The higher the level of abstraction be it in terms of type of intervention or in terms of geographical scope, the more complex the evaluation – and hence the evaluation design – is likely to be. The most complex evaluation designs are those that address in parallel or jointly several intervention measures and at different levels of spatial disaggregation.

4. Evaluation need to be timed appropriately. It is important that the function of the evaluation is specified at the outset and in common by those commissioning an evaluation study and those in charge of carrying it out. The function of the evaluation study is determined by its timing with regard to the phase in which the initiative to be evaluated is to be found. The function of the evaluation study influences the evaluation design.

5. Determine the time horizon of the evaluation and accordingly specify what type of effects will be looked at – outputs, results or outcomes. Subsequently determine the indicators to be used paying attention to time series and distribution effects.

6. A methodology mix with regard to data collection is advisable. The final decision must be justified with reference to the evaluation design but also considering pragmatic constraints. Where existing sources are used, a transformation of key variables might be unavoidable but must be documented.

7. The use of models and/or statistics to guide data analysis in evaluation must be informed by a good level of theoretical knowledge and relevant documentation. Assumptions made in the evaluation design regarding the subject of study must be consistent with those of models or statistical theory. Non-statistical methods can be used to complement statistical methods. Their merit lies primarily in their ability to capitalise on users' and stakeholders' interests and in their ability to communicate evaluation results in a policy-relevant and comprehensive way.

8. The assessment techniques to be used in evaluation must be chosen at the outset of the evaluation, thus influencing the latter's design. They should in principle inform data collection strategies or methods of analyses. A combination of impact assessment techniques with multi-criteria analysis or cost-benefit analysis is the typical approach in transport. Scenarios are indispensable when long-term effects are taken into the loop. Policy analysis as a form of meta-analysis can assist in the integration of results and the working out of strategic impacts.

9. In modern evaluation it is important for the evaluation team to be up-to-date with the state-of-the-art in the field of study. This will also allow it to determine the extent to which new research will be necessary for carrying out the evaluation or whether it is possible to rely on existing information. If new research is necessary this must be taken into account in the time plan and budget.

10. Determine the composition of the evaluation team after outlining the evaluation design. The latter should indicate what types of expertise are

necessary and guide the selection of team members and division of labour among them.

11. The attitude of commissioning agencies is an important facilitator or barrier to the successful implementation of an evaluation and the acceptability of its results. In commissioning agencies where there are no standardised procedures for planning and processing evaluations, evaluation results are of little use regardless of their quality.

12. When considering stakeholders interests in evaluation, it must be clear to the evaluation team and the agency commissioning the evaluation that this might require changes in the evaluation design and certainly changes in the communication style.

13. The evaluation design handbook and the final report are the two most important outputs of an evaluation. They should be written in a comprehensive manner and be explicit on the results and on underlying assumptions.

Good Evaluation Design Checklist GED-13

The Good Evaluation Design Checklist GED-16 proposed below has been developed on the basis of the insights to good evaluation practice described above. Using the GED-16 can help evaluation teams judge the complexity of the evaluation design required by the terms of reference of evaluation studies and commissioning agencies evaluated proposed evaluation designs or map terms of reference.

One may think of this checklist as the minimum list of indicators that need to be assessed prior to embarking on an evaluation. The answers to these questions should guide the evaluation design. T is used to refer to the initiative proposed for evaluation.

1) The purpose of the evaluation – why, what, for whom

Assess why the evaluation is being undertaken, for whom or for what purpose has it been initiated. Is it meant to address generally valid objectives; the specific original objectives of T or a mixture of the two?

Does the evaluation enjoy support from the top and is responding to effective demand in accordance with realistic expectations and linked with appropriate budget and time lines? If not how do you propose to deal with this problem?

Is the involvement of stakeholders a requirement of the evaluation study? If so, how do you propose to take stakeholders' interests into account and at what stage, i.e. at the stage of data collection, data analysis, formal assessment, or simply with regard to the communication of results?

Which stakeholder interests or world views you are not considering in your study and why not? Is this justified by the objective or scope of the evaluation and/or the choice of methods?

Will the results of your evaluation be made public? Is the report meant for dissemination within the general public? If so, have you taken adequate consideration of this in the writing of the final report?

Do you envisage the organisation of direct citizen participation procedures?

2) Scope of the evaluation in terms of the transport area targeted

Using the classification scheme proposed in this document, specify whether the subject of evaluation T is an infrastructure investment initiative, an urban network organisation initiative, a traffic system management programme, a regulation relating to technical or environmental standards, a regulation relating to market access or an economic measure. The type of measure under study influences the evaluation design and the method chosen. Specifying the scope of the evaluation helps also streamline the background information necessary for the study.

3) Geographical scope of the evaluation

Specify the geographical scope of the evaluation of T. Is this the national level and in turn the urban level, the inter-urban national level, the regional or inter-regional level; or is it the European level and in turn the regional cross-border, the inter-urban cross-border, the national cross-border or the trans-national network? As above, the geographical scope of the initiative under study influences the evaluation design and the method chosen. It also helps streamline the background information necessary for the study.

4) Stage of T in the decision process

What is the timing of the evaluation? Are we talking about an ex-ante, an intermediate or an ex-post evaluation study? Is the evaluation appropriately timed, i.e. is enough time allowed to process the evaluation results in the decision process? If not, reconsider the function of the evaluation with the commissioning agency.

5) Time horizon of the evaluation

What is the time-horizon of the evaluation? Short-term, mid-term, long-term or a mixture of the above? Is the time horizon proposed consistent with the objectives of

the evaluation (see indicator (1) above)? If not, re-consider the time horizon of the evaluation or the latter's objectives with the commissioning agency.

6) Types of effects to be studied

What types of effects are to be studied? Is the target the T outputs, i.e. the tangible effects of T in return for the public expenditure committed; the T results, i.e. the immediate short-term effects of T on a set of impact groups, or the T outcomes, i.e. the long-term effects of T on a set of impact groups or society as a whole.

Are the effects to be studied consistent with the objective and time horizon of the evaluation (see indicator 1 and 5 above)? If not re-discuss the effects to be studied or alternatively the objectives and time horizon of the evaluation with the commissioning agency.

7) Number of effects to be studied

Using the input from indicator (6) above, specify the number of effects to be studied. Accordingly you can obtain an idea on the number of indicators that will be necessary. Usually this should not be significantly smaller than the number of effects under study.

8) Availability / scope of context indicators

Using the input from indicators (2), (3) and (5) streamline the type of background information to be necessitated by the study. Carry out a preliminary assessment of the availability of relevant indicators or information (generally and to the evaluation team more specifically).

9) Methods in data collection

What methods for data collection do you propose to use? Specify these by using the classification scheme proposed in this handbook. Justify the choice and adequacy of these methods with regard to indicators (2), (3), (5) and (8) above. Revise accordingly.

Specify the time and resources necessitated by the choice of methods you are proposing and examine these against the budget and time line of the evaluation.

10) Methods in data analysis

What methods of data analysis do you propose to use? Specify these by using the classification scheme proposed in this handbook. Justify the choice and adequacy of these methods with regard to indicators (1), (2), (3), (5) and (8) above. Revise accordingly.

Are the methods you are proposing to use for data collection (indicator (9) above) going to deliver you the type of data you will need for analysis using the methods you propose to use?

Specify the time and resources necessitated by the choice of methods you are proposing and examine these against the budget and time line of the evaluation.

11) Methods for formal assessment

What methods of formal assessment do you propose to use? Specify these by using the classification scheme proposed in this handbook. Justify the choice and adequacy of these methods with regard to indicators (1), (2), (3), (5) and (8) above. Revise accordingly.

Are the methods you are proposing to use for data collection (indicator (9) above) and the methods you are proposing to use for data analysis (indicator (10) above) going to deliver you the data or type of analysis you will need for the methods of formal assessment you are proposing to use?

Specify the time and resources necessitated by the choice of methods you are proposing and examine these against the budget and time line of the evaluation.

12) Degree of reliance on existing databases, models or tools

Will you rely on existing databases or information sources for data collection?

Will you rely on existing models or tools for data analysis?

Will you rely on existing tools for assessment?

If the answer to either of the above questions is yes, then carry out a preliminary assessment as to whether the existing databases, models or tools you are proposing to use are (a) adequate for the evaluation of T; (b) entail assumptions that are not inconsistent with those you are making for the evaluation of T.

If the answer to either (a) or (b) above is to the negative but you are confident that the existing databases, models or tools you are proposing to rely on can be altered to fit your needs, examine whether the property and user rights governing the above are such that would allow you such revisions. If not you might need to re-consider their use or negotiate in advance with their owners.

13) Disciplinary make-up of evaluation team

What is the disciplinary make-up of the evaluation team you are proposing? Are the skills necessitated by the evaluation design as outlined by your answers to the above indicators (in particular the choice of methods – indicators (9) to (11) above)

reflected in the composition of the evaluation team. If not fully, how do you propose to make up for missing skills?

7 Annex C. TRANS-TALK Documentation

Besides the three workshops, TRANS-TALK has produced the following documentation:

The Theory and Practice of Evaluation; Conclusions for the First TRANS-TALK Workshop, July 2000 [**Deliverable 2**], Prepared by Liana Giorgi and Annuradha Tandon, ICCR

Evaluation Approaches and Database; Results of a Survey among 4FP Projects, May 2001 [**Deliverable 4**], Prepared by Dimitrios Tsamboulas, Costas Panou and Panayiotis Tsakiris, NTUA

Projects, Programmes, Policies: Evaluation Needs and Capabilities; Conclusions from the Second TRANS-TALK Workshop, August 2001 [**Deliverable 5**], Prepared by Alan Pearman, Peter Mackie and John Nellthorp, I.T.S. Leeds

The TRANS-TALK Final Report, With Conclusions from all Three TRANS-TALK Workshops and Integration of Findings, [**Deliverable 7**], Prepared by Liana Giorgi and Ronald J. Pohoryles, ICCR with input from all TRANS-TALK Partners

Guidelines for Improving Evaluation Practices in Transport, **A Consultation Document**, Prepared by Liana Giorgi, ICCR with input from all TRANS-TALK partners

The TRANS-TALK Virtual Library, www.iccr-international.org/trans-talk/library

Project and Policy Evaluation in Transport. **Book** edited by Giorgi, L. and Pearman, A. with the collaboration of Tandon, A., Tsamboulas, D. and Reynaud, C. and including contributions from First TRANS-TALK Workshop – Published by Ashgate, UK. Forthcoming in 2001

Transport Projects, Programmes and Policies; Evaluation Needs and Capabilities. **Book** edited by Pearman, A., Mackie, P. and Nellthorp, J. and including contributions from Second TRANS-TALK Workshop – Published by Ashgate, UK. Forthcoming in 2001.

8 References

Atkinson, R. and Cope, S. (1994), 'Urban Policy Evaluation: Science or Art?', Paper presented at the ESRC Research Seminar on Urban Policy Evaluation, Cardiff

Banister, D. and Berechman, J. (eds.) (1993), *Transport in a Unified Europe*, Amsterdam, North-Holland

Banister, D. and Lichfield, N. (1995), 'The Key Issues in Transport and Urban Development', in Banister, D. (ed.), *Transport and Urban Development*, London, E&FN Spon

Beuthe, M. (2000), 'Methods of Transport projects Evaluation: From Cost-Benefit to Multi-criteria and Decision Framework', Paper presented at the first TRANSTALK workshop, May 29-31 2000, Brussels, forthcoming as publication

Beuthe, M. *et al.* (2000), 'A Practical multicriteria methodology for assessing Public Investments', *Socio-Economic Planning Sciences*, No. 34, Elsevier

Bristow, A.L. and Nellthorp, J. (2000) Transport Project Appraisal in the European Union, *Transport Policy* vol. 7, p51-60.

Chelimsky, E. and Shahdish W.R. (eds.) (1997), *Evaluation for the 21st Century*, London, SAGE

CODE-TEN Final Report, (2000), *The DECODE Method – Theory and Application*, forthcoming as publication

Commission of the European Communities (1992), *Cost-benefit and Multi-Criteria Analysis for New Road Construction*, Euret Programme, Brussels

Cusco, V.P., DG TREN, EC (2000), 'Recent Development of the Common Transport Policy', Paper presented at the first TRANSTALK workshop, May 29-31, Brussels

Dasgupta, P. (1982), *The Control of Resources*, Great Britain, Harvard University Press.

Dodgson JS (1973), "External Effects and Secondary Benefits in Road Investment Appraisal" in *Journal of Transport Economics and Policy* 7(2).

Dodgson, J., Spackman, M., Pearman, A.D. and Phillips, L. D., *Multi-Criteria Analysis: a Manual*, Department of the Environment, Transport and the Regions, London, pp.158, 2000. ISBN: 1 85112 454 3. Also at: http://www.environment.detr.gov.uk/multicriteria/index.htm

Engel, U. and Pötschke, M. (1998), 'Willingness to Pay for the Environment: Social Structure, Value Orientations and Environmental Behaviour in a Multi-level Perspective', *Innovation*, Vol. 11, No. 3

European Commission (1992), *The future development of the common transport policy - A global approach to the construction of a Community framework for sustainable mobility*, COM(92) 494, Brussels

European Commission (1993), *European Transport Policy in the 90s*, Brussels

European Commission (1993), *The future development of the common transport policy. A global approach to the construction of a community framework for sustainable mobility.* Bulletin of the European Communities, Supplement 3/93

European Commission (1993), *Project Cycle Management – Integration Approach and Logical Framework – A Manual*, Brussels, EC-DGVIII

European Commission (1994), *Strategic Environmental Assessment – Existing Methodology*, Brussels, EC

European Commission (1995), *The Common Transport Policy Action Programme 1995-2000*, Brussels

European Commission (1995), *The Trans-European transport network; transforming patchwork into a network*, Brussels

European Commission (1995), 'Amended proposal for a European Parliament and Council Decision on the Community guidelines for the development of the trans-European transport network', *Official Journal C97*, Vol.38

European Commission (1995), *Common Guide for Monitoring and Interim Evaluations*, Brussels, EC Structural Funds

European Commission and Transport Evaluation Group (1995), *Framework for Transport Investment Evaluation*, Brussels, EC

European Commission (1996), APAS-Strategic Transport, *Cost-benefit and multi-criteria analysis for nodal centres for goods*, Brussels, EC

European Commission (1996), EURET-Concerted Action 1.1, *Cost-benefit and multi-criteria analysis for new road construction* , Brussels, EC

European Commission (1997), *Case Studies on Strategic Environmental Assessment; Final Report: Volumes 1 & 2*, Brussels, EC

European Commission (1997), *Financial and Economic Analysis of Development Projects*, Luxembourg, EC

European Commission, DGXIX, (1997), *Good Practice Evaluation Guidelines Handbook*, EC

European Commission (1997), *Guide to Cost-Benefit Analysis of Major Projects*, Brussels, DG-XVI (in the context of EC regional policy)

European Commission (1997), *Evaluating EU Expenditure Programmes; A Guide*, Brussels, DG-XIX-Budget

European Commission (1999), PHARE – *Transport Infrastructure Needs and Assessment*, TINA Secretariat

European Commission (1999), *MEANS collection– Evaluating Socio-economic Programmes*, Brussels, EC (five volumes)

European Commission (1999), Project Cycle Management Training Handbook, Brussels, EC SCR-Evaluation

European Commission (2000), *Good Practice Guidelines for the Management of the Evaluation Function*, Brussels, EC DG-Audit

European Commission (2000), *Evaluation of the Implementation and Results of SEM 2000*, Final Report, Brussels, EC

ECMT (1992), *Evaluating Investment in Transport Infrastructure*, Paris, ECMT

European Conference of Ministers of Transport (ECMT) (1998*), Efficient Transport for Europe: Policies for Internalisation of External Costs.* OECD, Paris.

Fischer, F. (1995), *Evaluating Public Policy*, Chicago, Nelson hall Publishers

Foster CD (2001), 'Michael Beesley and cost benefit analysis' in *Journal of Transport Economics and Policy,* 35, p3-30.

Freeman, H.E:, Rossi, P.H. and Wright, S.R. (1980), *Doing Evaluations*, OECD, Paris

Giorgi, L. and Pohoryles, R. (1998), *TENASSESS Deliverable 5, Interconnections Among Tasks: A Guide to the Fourth Framework Strategic Transport Research Programme*, Vienna / Brussels, ICCR / EC

Glaister, S. et al (1998), *Transport Policy in Britain*, London, MacMillian Press Ltd.

Goodwin, P. and Wright, G. (1998), *Decision Analysis for Management Judgment,* 2nd. edition, Wiley

Grant-Muller SM, Mackie PJ, Nellthorp J and Pearman AD (2001), 'Economic Appraisal of European Transport Projects - The State of the Art Revisited' in *Transport Reviews* 21(2), p237-261.

Gwilliam, K. (2000), 'Transport Project Appraisal at the World Bank', Paper presented at the first TRANSTALK workshop, May 29-31 2000, Brussels, forthcoming as publication

Haight, F. (2000), 'Evaluation of Projects and Programmes: Principles and Examples', Paper presented at the first TRANSTALK workshop, May 29-31, Brussels, forthcoming as publication

Hogwood, B. W. and Dunn, L. A. (1984), *Policy Analysis for the Real World*, Oxford, Oxford University Press.

Homeyer, H.von, (2000), 'Enlarging EU Environmental Policy-The Challenges of Flexibility and Integration', Paper presented at the first TRANSTALK workshop, May 29-31 2000, Brussels, forthcoming as publication

Hoon Oum, T. *et al.* (eds.) (1997), *Transport Economics; Selected Readings*, Harwood Academic Publishers

Jara-Diaz SR (1986), 'On the relation between users' benefits and the economic effects of transportation activities', *Journal of Regional Science* 26, p379-391.

Keeney, R.L. and Raiffa, H. (1976), *Decisions with Multiple Objectives,* Wiley.

Layard, R. and Glaister, S. (Edited) (1996), *Cost-Benefit Analysis*, Cambridge University Press

Leleur, S. (1995), *Road Infrastructure Planning*, Denmark, Polyteknisk

Mackie, P. Nellthorp, J. Pearman, A. (2000), 'Appraisal Issues for Trans-European Network Projects', Paper presented at the first TRANSTALK workshop, May 29-31 2000, Brussels, forthcoming as publication

Martinez F and Araya CA (1998), 'Land use impacts of transport projects: user benefits, rents and externalities'. Paper presented at the 8th World Conference on Transport Research, Antwerp.

McGowan, F. (2000), 'The EU Policy-Making Process', Paper presented at the first TRANSTALK workshop, May 29-31 2000, Brussels, forthcoming as publication

Mohring, H. and Williamson (1969) Scale and "Industrial Reorganisation" Economies of Transport Improvements, *Journal of Transport Economics and Policy,* vol 3*,* pp251-271.

Nellthorp J., Mackie P.J. and Bristow A.L. (1998), *Measurement and valuation of the impacts of transport initiatives,* Deliverable D9, (Restricted), EUNET Project - Socio Economic and Spatial Impacts of Transport (Contract: ST-96-SC.037), Institute for Transport Studies, University of Leeds.

Ney, S. (2000), 'Understanding Accessibility', in Giorgi, L. and Pohoryles R. J. (eds.), *Transport Policy and Research: What Future?*, Ashgate

OECD (1998), *Best Practice Guidelines for Evaluation*, PUMA Policy Brief No. 5, Paris, OECD

OECD (2001), *Outcome Focused Management and Budgeting*, Paris, OECD

OEEI (2000), *Appraisal of Infrastructural Projects: Guide for Cost-Benefit Analysis.*

Olson David L. (1996), *Decision Aids for Selection Problems*, New York, Springer

Palumbo, D.J., (ed.), (1987), *The Politics of Program Evaluation*, Sage, Newbury Park, Cal.

Parsons, W. (1995), *Public Policy: An Introduction to the Theory and Practice of Policy Analysis*, Aldershot, Edward Elgar

Parsons, W. (2000), 'Analytical Frameworks for Policy and Project Evaluation: From Welfare Economic and Public Choice to Management Approaches', Paper presented at the first TRANSTALK workshop, May 29-31 2000, Brussels, forthcoming as publication

Pearce, D.W. and Nash, C.A. (1981), *The Social Appraisal of Projects: a Text in Cost-Benefit Analysis,* London: Macmillan.

Pearce, D. Hett, T. *et al.*, (1999*), Review of Technical Guidance on Environmental Appraisal,* UK Department of the Environment, Transport and Regions (DETR), UK

Phipps L. *et al.* (1999), *Evaluating Environmental Schemes – A Catalyst to Accelerate Good Practice*, UK Evaluation Society, UK

Ponti, M. (1997), *TENASSESS Working Paper on Conflicts*, TRT

Rambow, G. (2000), 'The EU Enlargement and its impacts on European Policies', Paper presented at the first TRANSTALK workshop, May29-31, Brussels, forthcoming a publication

Rennings, K. (2000), 'Criteria for Evaluation towards Sustainability', Paper presented at the first TRANSTALK workshop, May 29-31 2000, Brussels, forthcoming as publication

Reynaux, C. and Panetta, A. (2001) Overview of Evaluation Experiences and Lessons learnt through Different Transport Planning Procedures in Europe, INRETS, Paris. (Report prepared as part of TRANS-TALK, work package 3).

Ritter, H.W.J and M.M. Webber. (1973). 'Dilemmas in a General Theory of Planning', *Policy Sciences,* Vol. 4, pp.155-169

Rossi, P.H and Freeman, H.E. (1993), *Evaluation : A Systematic Approach*, California, SAGE

Rosenhead, J. V. (ed.) (1989), *Rational Analysis for a Problematic World*, Oxford, Wiley

Roy, R (1995), *Lost and found: the community component of the economic return on the investment in PBKAL.* European Centre for Infrastructure Studies Report, Rotterdam.

Rus G. de, Nach C. (eds.) (1996), *Recent Developments in Transport Economics*, Hants, Avebury

SACTRA (Standing Advisory Committee on Trunk Road Assessment) (1999), *Transport and the Economy.* London: TSO.

Sansom, T., Pearman, A. D., Matthews, B. and Nellthorp, J. (1999), *The SITPRO Methodology*, Deliverable 2, SITPRO (Study of the Impacts of the Transport RTD Programme), Brussels, EC

Schweizerische Evaluationsgesellschaft (1999), *Evaluations-Standards*, (authors: Thomas Widmer, Charles Landert and Nicole Bachmann), SEVAL

Sen, A.K., (1997), 'Shadow Prices and Markets', in Layard, R. and Glaister, S. (eds.), *Cost-Benefit Analysis*, Cambridge, Cambridge University Press

Simmonds D and Marcial Echenique & Partners (1999), *Review of Land Use and Transport Interaction Models.* Report to SACTRA Committee. London: DETR.

Sugden, R. and Williams, A. (1978), *The Principles of Practical Cost-Benefit Analysis*, Oxford University Press

TINA Secretariat (1999), *Socio-Economic Cost-Benefit Analysis in the Context of Project Appraisals for Developing a Trans-European Network in Bulgaria, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Slovenia*. TINA 015/Final. TINA Secretariat: Vienna.

Tinch R (1996), *The Valuation of Environmental Externalities.* London, HMSO.

Tsamboulas, D. Mikroudis, G. (2000), 'EFECT – Evaluation Framework of Environmental Impacts and Costs of Transport Initiatives', *Transport Research Part D*, No. 5, 2000,

Turró, M. (1999), *Going Trans-European, Planning and Financing transport network for Europe*, Amsterdam, Pergamon

U.K. DETR (1999), *Local Evaluation for Regeneration Partnerships – Good Practice Guide*, London, HMSO

U.K. DETR (1999), *Proposal for a Good Practice Guide on Sustainable Appraisal of Regional Planning Guidance*, London, HMSO

U.K. DETR (1999), *Review of Technical Guidance on Environmental Appraisal*, (Pearce, D. et al.), UK

UK DETR (1999), *Transport and the Economy*, The Standing Advisory Committee on Trunk Road Assessment

U.K. DETR (2000), *Guide to Good Practice*, First Draft, London, DETR

U.K. DETR (2000), *Guidance on the Methodology for Multi-Modal Studies*, London, DETR

U.K. Treasury (1997), *The Green Book; Appraisal and Evaluation in Central Government*, Treasury Guidance, London, HMSO

UK Treasury (1997), *The Green Book – Appraisal and Evaluation in Central Government*, London, The Stationary Office

U.S. GAO (1990), *Case Study Evaluation*, GAO/PEMD-91-10-1.9, Washington, General Accounting Office, Programme Evaluation and Methodology Division

U.S. GAO (1990), *Prospective Evaluation Methods; The Prospective Evaluation Synthesis*, GAO/PEMD-10.1.10, Washington, General Accounting Office, Programme Evaluation and Methodology Division

U.S. GAO (1991), *Using Structured Interview Techniques*, GAO/PEMD-10.1.5, Washington, General Accounting Office, Programme Evaluation and Methodology Division

US GAO (1991), *Designing Evaluations,* Program Evaluation and Methodology Division, US

U.S. GAO (1992), *Quantitative Data Analysis: An Introduction*, GAO/PEMD-10.1.11

U.S. GAO (1998), *An Evaluator's Guide to Assessing Agencies Annual Performance Plans*, GAO/GGD-10.1.20, Washington, General Accounting Office

U.S. GAO (1998), *Performance Measurement and Evaluation; Definitions and Relationships*, GAO/GGD-98-26, Washington, General Accounting Office

U.S. GAO (1999), *Performance Budeting; Initial Experiences under the Results Act in Linking Plans with Budgets*, GAO/AIMD/GGD-99-67, Washington, General Accounting Office

U.S. GAO (2000), *Human Capital – A Self-Assessment Checklist for Agency Leaders*, GAO/OCG-00-14G, Washington, General Accounting Office

U.S. GAO (2000), *GAO's Strategic Plan Framework*, US

U.S. GAO (2001), *Internal Control Management and Evaluation Tool*, GAO-01-131G, Washington, General Accounting Office, Internal Control Division

U.S. Office Management and Budget (OMB) Circular A-94, (1992*), Guidelines and Discount Rates for Benefit-cost Analysis of Federal Programs* , US

Wallace, H. and Wallace, W. (1996), *Policy Making in the European Union*, New York, Oxford University Press

Williams HCWL (1977), 'On the formation of travel demand models and economic evaluation measures of user benefit', *Environment and Planning A* 9, p285.

Wilson A (1998), "Land-use/transport interaction models - Past and future" in *Journal of Transport Economics and Policy*, 32(3).

9 Contact Details

The Interdisciplinary Centre for Comparative Research
in the Social Sciences (ICCR),
Schottenfeldgasse 69/1,
A-1070 Vienna
Contact person: Liana Giorgi
Tel: (43-1) 524 13 93 160
Fax: (43-1) 524 13 93 200
E-mail: l.giorgi@iccr-international.org

Institute of Transport Studies,
The University of Leeds,
Woodhouse Lane,
UK-LS2 9JT Leeds
Contact person: Alan Pearman
Tel: (44-113) 233 4080
Fax: (44-113) 233 3906
E-mail: adp@lubs.leeds.ac.uk

INRETS,
2, avenue du General Malleret-Joinville,
FR-94114 Arcueil Cedex
Contact person: Christian Reynaud
Tel: (33-1) 47 407 262
Fax: (33-1) 45 475 606
E-mail: reynaud@inrets.fr

Faculty of Civil Engineering,
National Technical University of Athens,
5, Iroon,
Polytechniou Str.
Zografou Campus,
GR-157 73 Zografou Athens
Contact person: Dimitrios Tsamboulas
Tel: (30-1) 772 1367
Fax: (30-1) 772 1327
E-mail: dtsamb@central.ntual.gr

School of Economic and Social Studies,
University of East Anglia,
University Plain,
UK-NR4 7TJ Norwich
Contact person: Robert Sugden
Tel: (44-160) 359 3423
Fax: (44-160) 325 0434
E-mail: r.sugden@uea.ac.uk