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 thematic network on 'Policy and Project Evaluation Methodologies' 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, EC-REGIO, 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?

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