

Summary:

Evaluation traditions and criteria in the transport sector

This report discusses different evaluation types within the transport sector, in order to track important challenges that evaluators face.

Initially, we distinguish between different approaches of evaluation. Elvik's methodological approach of meta-evaluations emphasises that evaluations have to fulfil key methodological criteria such as statistical, theoretical, internal and external validity. The employer/principal approach emphasises that evaluations should improve knowledge of to what degree public policies leads to goal attainment, whereas Vedung's approach emphasises the importance of pluralism – pluralism in the sense that different approaches are adequate for different purposes. However, all three perspectives share an emphasis on the counterfactual as a key challenge in evaluation. An evaluation should answer, or at least address, the question of what would have happened if the measure introduced and evaluated had not been introduced. In addition, we introduce an approach rooted in the sociology of knowledge, emphasising that different evaluations are rooted in different disciplines, and based on different assumption, concepts and models that ultimately leads to different “findings”. Hence, this approach accentuates the question of theoretical comparability and compatibility between disciplines.

Several different evaluations are described and discussed in the report. In the first chapter measures introduced in order to improve public transport are discussed. The project evaluates measures introduced by principals. The local authorities (the principals) provide the data to the evaluators. Also, the choice of cases follows funding. The design gathers data from before and after the introduction of measures, a design that in principle should be able to produce data that would meet the criteria of statistical validity. In this evaluation, however, this requirement is only partially met due to lack of adequate data, in particular on the dependent variable which is passenger-growth. The theoretical framework is based on “middle-range” concepts (such as preferences, attitudes, options, actions) that may be related to more general theories in different disciplines (sociology, psychology), but they are merely presented as correlations. There is also a lack of internal and theoretical validity due to missing statistical validity. Also the external validity is not discussed. This evaluation is viewed as typical for multidimensional variable analyses - it is an evaluation with high potential for analysing change based on correlations between variables. Its external validity, has to be addressed either through theoretical or substantial comparison.

The principal approach's imperative is that evaluations should be made in cases of high significance or risk. Following this, the third chapter addresses evaluations of large-scale infrastructure projects. The chapter refers to research that find

miscalculations and cost overruns as a major problem in transport policy, both nationally and internationally. Two approaches try to explain this problem.

First, ex post cost benefit analyses are based on the assumption that these miscalculations may be reduced by improving the methods. Such analyses represent an exemption in our report in the sense that it is an evaluation anchored in a theoretical discipline (welfare economics), has clear assumptions, and has a design that – given its validity area – has a high potential for meeting statistical, theoretical, internal and external validity.

However, in the cases of ex post cost benefit analyses described in our report, the data provided to the evaluators was scarce. Due to lack of information of the ante situation, the evaluators were not able to make proper ex post cost benefit analyses. For that reason they could not make conclusions of miscalculations, but they contributed with advices on how to improve documentation and methods.

Our report points out that, although some aspects of the valuation of time, environment and accidents are continuously debated, tested and sophisticated within the discipline of economics, some aspects are merely assumptions – and hence not tested. On one hand, ex post cost benefit analyses are paradigmatic as a type of evaluation that meets the criteria – or ideal – of meta-evaluation. On the other hand, such analyses are based on assumptions that may be questioned. The valuation of time illustrates our argument. It is an important variable in the determination of the cost benefit ratio, but the assumption of aggregation of time values is controversial. Our report gives an overview of criticism against cost benefit analyses in general. It concludes that although some criticism is inconsistent, such interdisciplinary controversies show that there is uncertainty connected to the results. Hence, discussions on key assumptions of cost benefit analyses in the interdisciplinary and public debate is requested, not the least because cost benefit analyses hold an important position in the administrative process - yet less legitimacy in the political process.

The second approach to evaluation of large-scale infrastructure projects is the organisational approach. This is an approach based on the assumption that there are not only methodological, but also organisational explanations of miscalculations and cost overruns. The registration of data, the calculations of costs and benefits, as well as the presentation of results are made in a certain organisational and institutional setting, e.g. calculations may be carried out by organisations that may have economic interests in the results, and in an institutional environment with scarce public resources, and in which the actors compete on projects.

Our report presents and discusses the Danish pioneer research by Flyvbjerg in particular. His work aims at explaining why transport projects tend to be more costly than budgeted. The design of his evaluations is characterised by middle-range theoretical concepts and is interdisciplinary anchored. The design seems to meet statistical and internal validity, however, there is a lack of such internal validity in Flyvbjerg's work due to his interpretation of the data. Moreover, the theoretical validity in his work is problematic. Our report argues that some of his conclusions are inchoated, in particular when it comes to the notion of a 'fooling game' which results in 'the survival of the unfittest.' Also the conclusion that

projects that are the misrepresented (fooling the principal) will win, rather than the best projects, is problematic.

Moreover, our report finds that the approach lacks variance on the actor variable as well as the institution variable, e.g. Flyvbjerg assumes that all actors are equal (all actors are opportunistic, rather than e.g. also acting in accordance with scientific norms), and all institutional incentive structures promote lies. However, the very improvements of incentive structures that the researchers propose, such as increased transparency, local cofinancing, and lump-sum rather than earmarked financing, are characteristics that vary across time and space. Hence, such variation should be included in the analyses. Moreover, such analyses would be of significance for further public and interdisciplinary debate on cost overruns and miscalculations of large infrastructure investment projects.

The third chapter thoroughly discusses organisational evaluations. The importance of such evaluations is acknowledged, firstly because organisational change often is introduced as a measure in itself, and secondly, because public measures such as new regulations, new economic incentives, and new procedures are always transposed through organisations. This implies that organisational characteristics affect the outcome of a change. However, as argued in our report, organisational evaluations are difficult in cases where an organisation is an independent or mediating variable, but there is a lack of information of the dependent variable. In such cases the question of the counterfactual remains unsolved.

This is a challenge of the organisational evaluation, the process evaluation of transport plans in ten Norwegian cities (TP 10), also addressed in the third chapter. Its evaluation design is based on an analytical framework based on concepts such as actors, processes, and local projects. The case studies in this evaluation are thoroughly described, but both the internal and theoretical validity is difficult to examine because it is not made explicit in the analyses. However, the results seem to have been validated through feed-back from interviewed actors.

We argue that the general conclusion is based on a premise that is not tested. The main argument is that TP 10 had limited effects because other important elements in the institutional environment remained unchanged. Despite the institutional changes which were introduced in order to reduce e.g. road building incentives, the incentive structure still made it rational for local authorities to prioritise road building as it maximized economic external resources (resources from the state and not local funding). The argument following from this diagnosis was that the organisational change was not strong enough.

However, counteracting hypotheses, i.e. additional changes in incentive structure would not alter priorities, were not addressed. In our report we argue that the evaluator of TP 10's argument may be right, but it is not empirically confirmed.

To sum up, it is a common characteristic of the evaluations discussed in our report, that they barely discuss the external validity of findings. If this is a general pattern, it implies that findings from evaluations are not given the critical scrutiny and control that follows from systematic comparison and critique.

Furthermore, our report enhances two types of transparency problems: the transparency problem related to complex statistical analyses, in which critique must be formulated by peers, and the problem related to qualitative analyses in which the validation sometimes is carried out by the informants. The potential bias of such transparency problems are likely to increase in strong epistemic communities of evaluators, administration and practitioners, sharing common beliefs of a given political problem, and how it is to be solved, and having monopoly of definitions of and discussions on these problems.