Dare to learn, continue measuring

In search of frameworks and methods for the evaluation of system and transition policies

October 2022

Foreword

The report of the Effect Measurement Expert Working Group, led by Jules Theeuwes, was published in 2012. The committee was made up of Dutch evaluation experts and advised the Ministry of Economic Affairs and Climate Policy on how the instruments of Dutch innovation policy could best be evaluated and which methods to use. The committee produced an overview of the available evaluation methods and conducted discussions with the officers responsible for policy instruments to obtain information on, among other things, the objective, target group, data collection and operation of policy instruments. On this basis, an advisory report was drawn up on an evaluation approach.

The essence of what later became established as the Theeuwes approach is that policy evaluations are assessed by applying econometric methods that account for selectivity issues (biases) and exogenous influences. Ultimately, policy evaluations are about providing empirical insight into the size and nature of the effects of a policy intervention and whether they would have occurred without government involvement and the use of taxpayers' money.

For ten years, the recommendations of the Effect Measurement Expert Working Group have been the guiding principles for the evaluations conducted in the Ministry of Economic Affairs and Climate Policy as part of its business policy. The Theeuwes approach has significantly advanced evaluation practice, but it is not applicable to all policy interventions. The approach is not always appropriate, particularly in the case of system and transition policy aimed at, for example, sustainability and digitalisation. Among the reasons is the fact that there is not always a single instrument with a clearly defined goal that can be evaluated. Moreover, the availability of data on inputs, outputs and outcomes is typically limited and conclusions on effectiveness and efficiency are often difficult to draw in the shorter term. Hence there is a need for new, different evaluation methods that can be deployed alongside the Theeuwes approach for the evaluation of system and transition policies.

The Ministry of Economic Affairs and Climate Policy appointed a new expert committee at the end of 2020 to develop a vision of evaluation practice and advise on the development of methods for system and transition policy evaluations. This report provides an overview of the scientifically available evaluation methods. It considers both quantitative and qualitative methods (and combinations of both) that can be used to reveal the effectiveness of system and transition policies. The report focuses particularly on the role of monitoring in evaluation due to the regular need for long-term policy efforts, on the legitimacy of government interventions and on the solidity and validity of the evaluation methods. Finally, it discusses a number of specific cases in order to illustrate the recommendations and decision aid that are being proposed.

The committee consisted of the following members: Rob Aalbers (CPB), Koen Frenken (Utrecht University), Matthijs Janssen (Utrecht University), Peter van der Knaap (Ministry of Foreign Affairs), Carl Koopmans (VU University), Valéry Pattyn (Leiden University), Gusta Renes (PBL), Reinhilde Veugelers (KU Leuven) and Bas ter Weel (University of Amsterdam). The committee was chaired by Bart Verspagen (Maastricht University) until March 2021, when the chairmanship was taken over by Bas ter Weel. The final report was written by Bas ter Weel (SEO Economic Research), Matthijs Janssen (Utrecht University / Dialogic), Michiel Bijlsma (SEO Economic Research) and Pieter Jan de Boer (Dialogic).

The ministerial experts were: Theo Roelandt (Economic Affairs and Climate Policy), Henry van der Wiel (Economic Affairs and Climate Policy), Kim Hermans (Economic Affairs and Climate Policy), Geert Thijssen (Netherlands Enterprise Agency), Wouter Panneman (Economic Affairs and Climate Policy), Henk Massink (Agriculture, Nature and Food Quality), Marcel Seip (Netherlands Enterprise Agency),

Alexander Buitenhuis (Agriculture, Nature and Food Quality) and Thomas Niaounakis (Finance). The secretariat was conducted by Robert Schaap (Netherlands Enterprise Agency).

This report is a first step towards evaluating system and transition policies. With the evaluation of this type of policy, we are at the beginning of a research agenda that is intended to provide clarity about the frameworks, methods and techniques that can adequately determine the effectiveness and efficiency of policies and the relevant data requirement. With thanks to the members of the committee, responsible officers, departmental experts, the secretary and authors of this report, work has started on a new set of methods and techniques that can help in determining the effectiveness and efficiency of policy.

Bas ter Weel

Chairman of the System and Transition Policy Evaluation Methods Committee

Executive summary

With its system and transition policies (S/T policies), the Dutch government aims to improve innovation systems and achieve transitions towards, for instance, a sustainable and digital economy in the Netherlands. S/T policy is the collective name for integrated policy programmes aimed at achieving alignment through coordination activities between various policy instruments in order to improve innovation systems and achieve transitions. These policies are legitimised by the presence of different forms of failure, such as external effects and information and coordination deficiencies. In this regard it is important that the costs of government intervention do not exceed the costs of failure. Although in many cases only a limited budget is available for S/T policies themselves, the policies apply to a comprehensive set of policy measures for which there is a substantial budget. Policies also have a large reach, in the sense that many economic actors and processes come into contact with them. It is therefore important to measure the effectiveness and efficiency of S/T policies as accurately as possible.

Assignment

The report of the Effect Measurement Expert Working Group (Theeuwes Committee) was published in 2012. The working group advised the Ministry of Economic Affairs and Climate Policy on how the instruments of Dutch innovation policy could best be evaluated and which methods to use. The essence of the advisory report is that a number of econometric methods should be applied in evaluations that require effect measurement. It turns out in practice, however, that econometric methods cannot be applied to all types of policy interventions. Particularly in the case of S/T policies, there is a lack of clear and validated methods and the academic literature does not yet provide any clear evaluation frameworks. In order to ascertain the effectiveness and efficiency of such policies, it is therefore necessary to find new, different evaluation methods to supplement those proposed by the Effect Measurement Expert Group.

The task assigned to the System and Transition Policy Evaluation Methods committee is as follows: To develop a vision for the evaluation practice and strategy of the Ministry of Economic Affairs and Climate Policy. In so doing, the committee focuses on advising on the development of methods for the evaluation of S/T policies. The advisory report provides an overview of the scientific literature on a range of policy issues in the innovation policy of the Ministry of Economic Affairs and Climate Policy. This concerns both quantitative and qualitative research methods that can reveal the effectiveness of system and transition policies. Attention is also devoted to the role of monitoring in evaluation, to the legitimacy of government interventions and to the evidential value of the evaluation methodology for system and transition policies.

This report only partly remedies the lack of evaluation methods, because we are at the beginning of a research agenda that aims to provide clarity about which frameworks, methods and techniques can adequately determine the effectiveness and efficiency of S/T policies and the relevant data requirement. There is no international standard for the evaluation of S/T policies and the development of scientific methods is in its infancy. Whereas an extensive literature with methods and techniques is available for instrument evaluations, there is no such literature for S/T policies. Furthermore, scarcely any data are available for the evaluation of policies, there is no counterfactual, and such policies are by their nature dynamic. Taking into account such challenges, this report explains what evaluation factors and methods are appropriate in the evaluation of S/T policies.

System and transition policies

S/T policies can be defined as influencing the production, consumption and distribution public goods and services, where these goods and services are the result of an interactive process between different actors, such as consumers, businesses, universities and government. These are policies that seek to bring about improvements with a coherent set of policy instruments to all of the production factors in order to achieve a particular societal objective. System policy focuses on improving the characteristics of the system as a whole, rather than on isolated sub-aspects of that system. Transition policy is concerned with adjusting the various elements that constitute a system (e.g. if new objectives have to be achieved or the system needs to operate in a different way). In the case of both system policy and transition policy, there is usually an experimental and learning process in which the pathway from problem to solution is not clear at the outset and even the solution sometimes cannot be clearly defined in advance. This means the policy must be designed adaptively, in terms of both its introduction and abolition, because the series of incentives and laws and regulations that are issued or abolished must be adjusted when the system encounters new issues.

Evaluation framework

Due to the lack of an international scientific evaluation standard, this report first adopts an evaluation framework based on a synthesis of the available scientific literature and existing evaluation practice. This framework comprises six perspectives: intervention logic, governance processes and policy mix, match between policy/policy mix and system bottlenecks, system strengthening/transformation, structural change and societal impact. For each specific perspective it is possible to draw up detailed hypotheses indicating what change can be expected. Each perspective in the evaluation framework highlights a different aspect of the operation and effectiveness of S/T policies. The perspectives on system and structural change and societal impact lend themselves to summative analyses with a focus on accountability, whereas in the perspectives on intervention rationale, governance processes and the policy mix and the alignment with issues the emphasis is on formative research into the underpinning and implementation of the policy.

In order to apply the evaluation framework perspectives in practice, it is important to operationalise them by drawing up a policy theory. In this regard the framework below provides some examples of questions to be answered.

Policy theory perspective	**Examples of questions**
Intervention logic	What are the types of failure that require S/T policies (i.e. with a focus on coordination and streamlining)? Where precisely are the issues, and what is the evidence for them? For example, without the policy, are there insufficient possibilities and policy options to contribute to system change and/or transitions, and why are the market and cooperating actors unable to find a solution themselves?
Governance and policy mix	Are there principles indicating what the coordination in the S/T policies should focus on, and does the coordination comply with those principles? Does the coordination lead to appropriate adjustments in the policy mix (i.e. modifications that align system changes and resulting activities better with policy objectives)? To what extent is this reflected in the development or phasing out of schemes and in adjustments to the financing and design of schemes?
Match between policy/policy mix and system bottlenecks	What are the issues (in terms of system processes or structures) that complicate system change or make transition processes more difficult, and how are S/T programmes and the policy instruments to which that policy applies responding to these? Is most of the energy focused on the major issues, and are results achieved in terms of strengthening the weakest system aspects?
System strengthening and transformation	Do we see improvements in the system processes that enable the development and dissemination of innovations? Is there more development/exchange of knowledge, market formation, etc. (depending on which process was weak) around desirable innovations?

Structural change	Do innovation systems / socio-technical / production-consumption systems change structurally? This concerns the transformation of entire systems, including the parts that fall outside the narrow scope of the schemes associated with the S/T policies. An initial indication would be whether they show more innovation activity.
Societal impact	Are the interim goals and ultimate objectives achieved? And to what extent is that the result of system and other changes attributable to the influence of the S/T policies? For example, if there are sectors in which CO_2 savings have been achieved, is that also due to innovations that have been demonstrably boosted by the policy?

Research methods

Once the evaluation framework has helped determine which policy aspects should be examined, the next question is which research methods are most appropriate. The diagram below provides an overview of methods and techniques that are appropriate for evaluating the effectiveness and efficiency of the different perspectives of S/T policies.

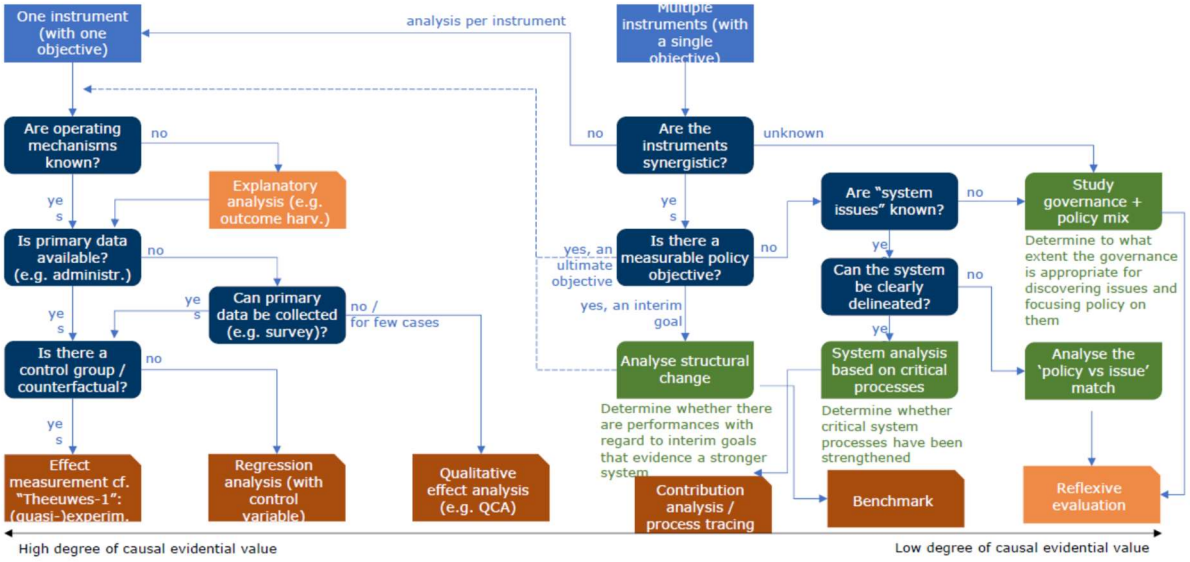
	Intervention logic	Governance and policy mix	Match between policy/policy mix and issues	System strengthening and transformation	Structural change	Societal impact
Effectiveness						
Systematic reviews & meta-analysis	Systematic reviews concern the identification of frameworks and principles to assess policy implementation. A meta-analysis is a numerical analysis.					
Reflexive evaluation	Explaining and (possibly jointly) interpreting mechanisms, starting from policy theory and implementation					
Case studies			Explaining mechanisms & determining outcomes			
Outcome harvesting			Explaining mechanisms, starting from outcomes (how is that outcome arrived at; what is the role of policy?)			
Contribution analysis & process tracing		Determining causality, starting from policy theory (gathering indications for the occurrence of a chain of outcomes)				
Qualitative Comparative Analysis (QCA)			Determining causality, starting from outcomes (what combination of factors determines variation in success?)			
Regression analysis				Measuring the size of effect, correcting for control variables		
Experimental/quasi-experimental methods				Determining causality & Measuring size of effect		
Efficiency						
Monitoring		Keeping track of developments, in predefined indicators				
Productivity examination (DEA, SFA, benchmark)				Comparing costs and/or performances (or the relationship between them)		
Simulations*					Estimating outcomes of complex interactions (including policy interactions)	
Societal cost-benefit analysis (SCBA)						SCBA

*Simulations can also indicate effectiveness, but they are included here only once for the sake of clarity

Decision aid

In order to help evaluators to select methods, a decision aid has been drawn up in the form of a flowchart. By going through this chart from top to bottom, evaluators can see the empirical methods that appear most suitable for the evaluation of a particular case. On the left-hand side of the chart are the methods with the highest causal evidential value. The report of the Effect Measurement Expert

Working Group has already described the specific techniques that can be considered when using experimental or quasi-experimental methods. Although the axis is not completely ordinal, the causal evidential value of methods decreases the further they are to the right in the figure. The methods to be considered are shown in the dark orange blocks. The method on the farthest right, reflexive evaluation, does not lend itself to statements about causal effects. This method block therefore has the same lighter colour as the 'explanatory analysis' block. The green blocks refer to the analysis frameworks that are relevant depending on the characteristics of the policy to be evaluated and the available information on it. They are linked to the perspectives in the evaluation framework.



Four examples

The usefulness of the frameworks, methods and the decision aid discussed in this report are applied to four cases that differ in the degree and manner in which they are examples of S/T policy. The variation on this point allows an exploration of the opportunities and limitations of the tools provided. The four selected cases are: The Mission-oriented Top Sector and Innovation Policy, the CO_2 reduction policy/climate policy, the Dutch Digitalisation Strategy and the Technology Pact.

Contents

1. Background

The report of the Effect Measurement Expert Working Group (Theeuwes Committee) was published in 2012. The working group advised the Ministry of Economic Affairs and Climate Policy on how the instruments of Dutch innovation policy could best be evaluated and which methods to use. The essence of the advisory report is the application of econometric methods in evaluations that require effect measurement. Such policy evaluations are about providing empirical insight into the size and nature of the effects of a policy intervention and whether they would have been achieved without government involvement and the use of taxpayers' money. Over the past decade, the recommendations of this working group have been the guiding principles for evaluations and the advice has helped advance evaluation practice in the Ministry of Economic Affairs and Climate Policy (and beyond).

It turns out in practice, however, that econometric methods cannot be applied to all types of policy interventions. There is a lack of clear methods particularly with regard to system and transition policies. New, different evaluation methods must therefore be sought in order to ascertain the effectiveness and efficiency of such policies. These methods will supplement those proposed by the Effect Measurement Expert Working Group.

A committee has been established by the Ministry of Economic Affairs and Climate Policy to advise on evaluation methods for system and transition policies. This committee is broadly composed with experts from scientific research and ministerial experts (see Foreword). The committee performed its work in the period from November 2000 to January 2022. This report is the result of that work.

1.1. Evaluation practice

The starting point is that an international evaluation standard and quantitative methods are already available for instrument evaluations (with one intervention, one user type and one clearly defined and measurable objective). This evaluation standard and these quantitative methods should be applied in accordance with the advice of the Effect Measurement Expert Working Group. However, there is no evaluation standard for system evaluations and the assessment of transition policies and the development of methods is in its infancy. Where possible, policy instruments that are part of a system or a transition should be evaluated as an instrument in accordance with the advice of the Effect Measurement Expert Working Group. Other methods are required in the case of the system or the transition itself.

The quality of evaluations depends greatly on the availability and quality of microdata files containing information on the response of businesses and individuals to the policy. There is a need to further improve the data infrastructure in the Netherlands, for example by investing in this data infrastructure prior to the introduction of new policies. This should ensure timely and adequate data collection on the input, output, outcome and impact of policy that, as far possible, is aligned with and can be linked to existing microdata files. Steps have been taken with regard to policy instruments over the past ten years and these must be continued and where necessary intensified. There is no significant data infrastructure for the evaluation of system and transition policies. This must be developed.

Most policy evaluations conducted in the past decade have only drawn conclusions about the direct effectiveness of policy instruments. The societal impact of policies (including external effects) and the effectiveness of policy can only rarely be assessed in a policy evaluation. This means that in many policy evaluations no clear and concise conclusions can be drawn about the effectiveness and efficiency of policies, even though there is a need for this on the part of the client (and in society). This

is partly due to deficient data, partly due to the timing of the evaluation, which is often too early, and partly due to the lack of a solid set of instruments for these 'higher-order' effects.

The evaluation system is aimed at assessing policy instruments to determine whether or not they have performed. In the Netherlands there is no culture of policy evaluations being used to learn in policy development. As a result, there is pressure on openness and transparency about what is going well and what could be improved, the possibility of interim adjustments and the desire to evaluate. In system and transition policies, where adjustments and experimentation are of great importance, this is a worrying development.

1.2. Requirement

The Ministry of Economic Affairs and Climate Policy needs to expand its evaluation toolbox. In particular, there is a need for insight into methods that set a standard for system and transition policies. The reasons for this requirement are as follows.

Transition policy, such as mission-oriented innovation policy or policy aimed at making energy-intensive manufacturing more sustainable, changes the policy perspective by formulating objectives for a goal to be achieved in the medium or long term. In the short term it is only possible to measure the direct output of a policy intervention towards the realisation of the goal. In addition, the realisation of the goal can be formulated in terms of a measurable objective, but the policy for the realisation of the goal is highly uncertain. The reason is that the most successful and optimum policy for the realisation of a goal is not predetermined. This means, for example, that although it is certain that new key technologies will radically change the world and will help to achieve a societal task, the possible applications of those technologies are not yet known, let alone already achieved in practice. In many cases, the technology must be further developed and adjusted in order to arrive at a concrete and usable application.

For the achievement of an objective various actors are associated with each other in a system and policy instruments influence each other. There is a high degree of dependence between the actors and policy instruments. In the short term, policy instruments can only be assessed in terms of the degree to which direct additional output is generated and what that contributes to the achievement of the objective. Monitoring the direct additional output and developing measurable indicators that say something about the direction of the observed development are important for the interim monitoring and adjustment of the policy. *Ex ante* evaluation appears less suitable for this, because the system already exists and is in operation. *Ex post* evaluation is also often impossible because cause and effect are not always clear in the system. Monitoring does appear to be applicable here, however. The question is therefore how the effectiveness and efficiency of system policy can be ascertained.

1.3. Assignment

The task assigned to the committee is as follows. To develop a vision for the evaluation practice and strategy of the Ministry of Economic Affairs and Climate Policy. This committee focuses on advice on the development of methods for the evaluation of system and transition policies. The advisory report provides an overview of the scientific literature on a range of policy issues in the innovation policy of the Ministry of Economic Affairs and Climate Policy. This concerns both quantitative and qualitative research methods that can reveal the effectiveness of system and transition policies. Attention is also drawn to the role of monitoring in evaluation, to the legitimacy of government intervention (market failure, system failure and transition failure) and to the place of the evaluation methodology for system and transition policy on the ‘effects ladder’ (see next section).

2. The committee's starting point

There have been two prominent trends in innovation policy over the past decade. First, due to the preference for evidence-based policy, there is growing interest in thorough econometric effect measurements, according to standards that are high on effect ladders like the Maryland Scientific Methods Scale.[1] This preference can be addressed with a quantitative evaluation of policy instruments, as elaborated by the Effect Measurement Expert Working Group. There is a crystallised scientific literature for the evaluation of individual instruments (with one intervention, one user type and one clearly defined and measurable goal), there is a recognised international evaluation standard, quantitative methods are available and data and a data infrastructure are often also available. Second, Europe in particular is seeing a rise in system and transition policies (S/T policies). These are policies that seek to bring about improvement by means of a coherent set of policy instruments to all of the production factors (the 'system') in order to achieve a particular societal objective.[2] Finally, there is an increase in innovation policies inspired by transition thinking. In this policy, which is often aimed at changing systems, the societal objective usually lies in the distant future and it is not clear what the most effective and efficient path to it will be. The policy goals (often intermediate goals set to achieve the ultimate societal objective) and approach can also change over time as new information becomes available.

2.1. Lack of an international evaluation standard

There is no international standard for the evaluation of S/T policies and the development of scientific methods is in its infancy. This report focuses on the development of guidelines for the evaluation of system and transition policies. These policy evaluations are shown in the two middle columns of the table below (Table 2.1). For the sake of completeness, this table also includes a column for the evaluation of institutions safeguarding public interests in the field of innovation, such as the professional research organisations known as TO2 institutions.[3] In addition to the periodic evaluation of the policy itself, the addition of new methods to the evaluation of practice is also useful for improving policy analyses. By expanding the evaluation practice, it will probably be possible to evaluate more parts of budget articles than is currently the case.

Aside from differences in the objects of S/T policy evaluation, differences may also arise in the evaluation criteria that are applied. This report focuses on how the criteria of effectiveness and efficiency are to be evaluated in the case of S/T policies. It also considers sub-aspects of effectiveness and efficiency, such as coherence, consistency and implementation of policy.

One of the aspects that determines the evaluation of effectiveness and efficiency is the type of data that can be used for evaluation. The quality of evaluations depends greatly on the availability of data that measure input, output, outcome and impact of policy and policy instruments. An evaluation of business policies by the Ministry of Economic Affairs (2020), an inventory of policy reviews in the pre-2011 period (Koopmans et al., 2019) and the policy evaluations Toolbox from 2020 reveal a need for further improvement of the data infrastructure, including by investing in the associated data collection and data development for monitoring and evaluation prior to introduction or at an early stage of new policies. This data infrastructure must be aligned with the existing infrastructure of microdata files

[1] Researchers dealing with evidence-based economic and innovation policy only consider to a limited extent *how* the precise design of an instrument relates to the generated impact, although it is recognised that in practice this can be just as important as knowing *whether* the policy has an impact (Hünermund & Czarnitzki, 2019; Duflo, 2017).

[2] The policy that focuses on this coherent package, for example by seeking greater cohesion, should itself also be viewed as a policy instrument.

[3] The TO2 institutions are five applied research organisations: Deltares, MARIN, the Royal Netherlands Aerospace Centre, TNO and Wageningen Research.

available from Statistics Netherlands. A lack of usable data is one of the reasons why only part of the policy can be assessed in a scientifically sound manner for effectiveness and an even smaller part for efficiency.

Table 2.1.Evaluation approach for policy issues in S/T policies

	Instrument evaluations	System policy evaluations	Transition policy evaluations	Evaluation of public institutes
Point of action	Actors	Operation & functions of innovation system	System and structural change	Institutions and framework conditions
Role of government	Market failure	System failure	Transition failure	Statutory tasks
Primary objective	Behavioural change	Improved operation of innovation system	Realisation of missions and societal objectives	Fulfilment of public duties
Primary form of intervention	Financial stimulus	Interaction and cohesion within the system	Societal objective and standards	Legislation, grants and basic subsidies
Evaluation approach	Natural and societal experiments	Determining operation of innovation system	Monitoring direction and status of development paths towards realisation	Achievements of goals and benchmarking
Reference	Aimed at influencing behaviour of individual actor	Aimed at systems and clusters	Aimed at systems and clusters	Aimed at organisations and legislation
Examples	WBSO, Innovation Credit	Top Sector policy	Mission-oriented innovation policy	TO2 institutions, universities, independent administrative bodies
Guidelines	Effect Measurement Expert Working Group (2012)	**This report**		Standard Evaluation Protocol

Another limitation is the availability of methods and techniques to evaluate policy. The methods and techniques available in the scientific literature mean that in most areas it is only possible to determine first-order outcomes of policy measures in terms of target range and effectiveness. The ultimate outcome and impact often remain outside the picture.[4] In addition, new data are needed to measure the efficiency of the policy and it appears necessary to develop new methods/models. Having reliable estimates of second- and third-order effects, including positive and negative externalities of policy instruments, is a necessary precondition for a better understanding of effectiveness and efficiency and for the ability to draw appropriate conclusions.

The above gaps apply to the evaluation of various types of policy. The reason for this report is that S/T policies have specific characteristics that make evaluation even more difficult. There are often even more data gaps, there are no counterfactuals as yet and there are no clear ways of evaluating these dynamic policies.

- *Data*. S/T policies focus on achieving effective and efficient processes, regulations and coordination to improve a system of actors and processes (e.g. the innovation system or the energy system) or on a transition to a new and sometimes unknown objective (e.g. further digitisation of the economy), so the availability of input, output, outcome and impact data is often

[4] An example is the effect of a subsidy scheme on the amount of research that businesses conduct. The impact on the amount of R&D is then measurable, but not the effects on economic growth and productivity. In some cases, macroeconomic equilibrium models can help to determine the economic impact of first-order effects.

limited. There is a need for data to monitor these processes and the degree of coordination, as a basis for analyses showing the extent to which a transition has been achieved or the extent to which the path towards a transition goal is being followed effectively and efficiently. Before conducting a policy evaluation, it is important that the client considers which data can and should be collected to measure effectiveness and efficiency.

- *Counterfactual*. Second, there is no counterfactual that measures the development of the system or the speed of a transition in the absence of S/T policies.[5] Not only is there often no second comparable system available, but S/T policies are also often a (possibly limited) part of a broader range of policy instruments, each with separate effects, or S/T policies consist of a combination of policy instruments. S/T policies are often not aimed at traditional policy instruments, but at coordination activities to direct existing policy instruments towards the desired system change or transition processes. As such, coordination can of course also be seen as an instrument. Nevertheless, in such cases it is difficult to attribute observed outcomes to a heterogeneous set of direct and indirect interventions. The aim of S/T policies is to improve the operation of these instruments by means of coordination. Isolating the effect of more coordination from the effect of the individual elements ('is it greater than the sum of parts?') is challenging. Due to the lack of the counterfactual, no clear and concise conclusions can be drawn about the effectiveness and efficiency of the policy. Counterfactuals are nevertheless sometimes available on a smaller scale, for example if the S/T policy includes a clearly delineated policy measure or instrument or if the coordination activity itself can be highlighted as an instrument to be evaluated. In that case it is possible to evaluate without having to make any statements at the system or transition level.
- *Dynamic instead of static.* Third, in S/T policies it is often clear what the ultimate objective is and how the market is currently failing, but it is unclear precisely which impediments will prevent the objective being achieved in the future when new and as yet unknown solutions are available. In that case a relevant question is whether existing policy involves processes and structures to highlight impediments, respond appropriately and adjust the policy. An important point is therefore whether the policy processes and structures are set up in such a way that they evolve along with the dynamic nature of systems and transitions. S/T policies do not only address known problems (such as the absence of system factors or transition processes) but also identify and mitigate emerging ones. That identification can relate, for example, to the 'weak' functions in the system that need to change in order to transform the system for the achievement of the objectives. Targeted actions can be taken to address individual issues. The instruments themselves may comprise relatively well-known and proven interventions, or policy approaches that gradually reveal which (possibly experimental) interventions work and why.

2.2. Scope of the advisory report

In contrast to the decades of research on evaluating economic and other policies based on a single issue and a single instrument, in the case of S/T policies there is no extensive evaluation tradition that provides ready-made tools for evaluating such policies. There is scientific literature on policy-based strengthening or transformation of innovation systems, production and consumption systems and socio-technical systems. That literature describes these different systems, but does not question the extent to which system changes have been brought about by policy interventions (or how that can be determined at all) and has no tradition of evaluating S/T policies.[6] This literature also focuses primarily

[5] A before/after measurement or a comparison with other countries or sectors can be helpful, but hardly ever constitutes a counterfactual due to the multitude of other factors that lead to differences.

[6] One explanation for the limited focus on evaluation of S/T policies is that these sometimes stem from perspectives on government policy that deviate from the traditional legitimate performance of government and are actually based on the roles of a networking and responsive government participating in 'intertwined dynamics' (Ministry of Economic Affairs, 2017).

on describing the overall system and less on the tipping points of a transition. There is literature on the evaluation of complex policy instruments based on the understanding that "the policy evaluation landscape is shifting; government is evaluating more varied policies, often made up of multiple projects, delivered in collaboration with many stakeholders, at multiple governance levels – central, regional and local" (Barbrook-Johnson et al., 2021). This literature often focuses on understanding policy (by involving policymakers themselves) and not on its effect or efficiency.

This report only partly remedies the lack of evaluation tools, because we are at the beginning of a research agenda that aims to provide clarity about which frameworks, methods and techniques can adequately determine the effectiveness and efficiency of S/T policies and the relevant data requirement. The analysis therefore explains what evaluation factors and methods are appropriate in the evaluation of S/T policies. In this regard we adhere to the adage that every policy evaluation is tailor-made (Van der Knaap et al., 2020, p. 5).

To obtain an overview, we first define what system and transition policies are and what the challenges are when it comes to evaluating them (Chapter 3). Chapter 4 presents perspectives and frameworks that describe how and on what basis system and transition policies can be evaluated. It also considers the operationalisation of these frameworks, in terms of questions and indicators. Chapter 5 answers the question of which empirical methods should be used to investigate those questions and indicators. Chapter 6 summarises the previous chapters in a step-by-step plan that seeks to offer concrete tools to prepare and design the evaluation of S/T policies. The emphasis here is on systematically assessing what is known about the policy intentions, how these will be driven forward and what analyses can be made. Chapter 7 then provides a decision aid for selecting the empirical method(s) with which to conduct the actual evaluation study. This takes concrete form in Chapter 8 with reflections on a number of specific cases relevant to the policy area of the Ministry of Economic Affairs and Climate Policy. Chapter 9 ends with an epilogue, which links the perspectives, frameworks and advice presented in this report to the steps to be taken in developing and applying an evaluation approach for S/T policies.

The alternative roles are associated with evaluation questions of a different kind, with less focus on accountability and more focus on highlighting how policy is implemented, which roles various parties play and how this can be improved. In this report we maintain a so-called accountability perspective on evaluation.

3. System and transition policies

This chapter defines what S/T policy is in the context of the Ministry of Economic Affairs and Climate Policy and how it operates. It thus specifies which type of policy falls within the scope of this report and how the policy is put together (Section 3.1), which types of failure are relevant and justify policy intervention (3.2), how the determination of effectiveness (3.3) and efficiency (3.4) differs from instrument evaluations, and what challenges this poses for the evaluation of S/T policies (3.5).

3.1. What are system and transition policies?

System and transition policies are about adapting structures, such as networks of actors or production-consumption systems, with the aim that those structures will generate socially desirable outcomes. This concerns the organisation of meso-level changes that are not brought about by influencing the behaviour of a single type of actor (including at the micro level). Instruments can of course be deployed at the micro level to encourage socially desirable behavioural change among consumers and producers.

In economic terms, S/T policies can be seen as influencing the production, consumption and distribution of public goods and services, where these goods and services are the result of an interactive process between different actors, such as consumers, businesses, universities and government. This result follows from an experimental and learning process in which the pathway from problem to solution is not clear at the outset and even the solution sometimes cannot be clearly defined in advance.

System-based thinking is valuable because it considers the overall process of producing, distributing and using public goods and services, as well as the impact of this process on economic and societal developments/transitions (Lundvall, 1992). With this approach, all activities in an economic system, as well as the organisations involved in these processes, can be seen as parts of a public purpose. The system contains feedback loops and multiple relationships between the organisations of which it is composed. Every element of the system is therefore important for the process as a whole and influences and is influenced by the other elements.

There are various ways in which the system approach can support and direct policy. First, it shifts the focus of policy from individual instruments to a set of instruments and the interactions between them. In some interactions, it is possible to use different instruments dealing with the production of public goods and services. The second way in which the systems approach is useful is that it focuses attention on scientific and technological inputs as well as on the processes that are important for the creation of public goods and services. A third way in which the system approach is valuable is that the behaviour of economic factors such as consumers and businesses but also public organisations is influenced by a range of instruments. These instruments create incentives to apply the results of scientific research and the use of experiments. Informal rules, standards, practices and routines are also important, such as organisations dealing with certifications or patents. Fourth, the system approach emphasises that policy can be applied at different levels, such as supranational (climate policy), national (top sector policy), regional (technology pact), sectoral and technological (digitisation). Finally, the focus of the analysis changes from the internal operation of an economic system to the way in which the system interacts with the outside world (for example climate policy).

If we take the innovation domain, we can distinguish three types of policy on which evaluations can focus: individual schemes such as the Research and Development Promotion Act (WBSO) (on which the report of the Effect Measurement Expert Working Group focuses), individual institutions such as the TO2 institutes (to be evaluated on the basis of the Standard Evaluation Protocol, for example) and

programmes. Programmes are overarching strategies in which several policy incentives are focused coherently on a specific objective. Such coherence is essential in the case of objectives that cannot be achieved with a single direct impulse but require the prior establishment of a system or transition process (involving multiple impulses) that supports the ultimate long-term objective. S/T policies fall in the latter category.

A characteristic of S/T policies is therefore that they concern integrated programmes which, through coordination activities, aim to achieve coordination between a series of (often disparate) policy incentives and instruments. What the policies ultimately aim to formulate is a medium-term policy objective or mission, but the path towards it and the right mix of policy incentives and instruments is uncertain. For example, it is unclear which new breakthrough technologies are relevant, how coordination (including regulation and incentives) can best be designed, or which policies and incentives should be abolished. A process of learning and testing is therefore required in order to achieve concrete and practicable fulfilment of the ultimate policy. This process is accompanied by experiments and pilots to find out what works and what does not, adjust incentives and rules, halt anything that fails prematurely or is inefficient, etc. Finally, it is important for the realisation of goals and transitions that actors are connected to each other in networks.

A *system policy* is deemed to exist if the policy focuses on characteristics of the system as a whole (e.g. an interrelated set of system elements) rather than on isolated sub-aspects of a system (e.g. the optimisation of a single element). For example, an energy tax is not a system policy because it seeks to influence the behaviour of individual actors through uniform price incentives. However, the instrument may be part of a system that uses various instruments to price the external effects of the use of fossil energy. There is deemed to be a *transition policy* if adjusting factors can transform systems (e.g. if new objectives have to be achieved or the system needs to operate in a different way). This means the policy must be designed adaptively, because the series of incentives that needs to be reinforced must be adjusted when the system encounters new issues or when new information becomes available.

3.2. Market failure, system failure and transition failure: possible new reasons for policies

In the scientific literature, market failure is one of the reasons for government intervention, which must be weighed against forms of government failure such as crowding out, costs of policy (such as permits and taxes), influence on policy by lobby groups or the chance that the government will (possibly unexpectedly) change policy. Different policy instruments seek to mitigate market failure. Just as markets can fail, there may be system failure and transition failure. In the case of innovation policy, for example, as well as boosting R&D expenditure, the goals of innovation policy include strengthening innovation systems and developing solutions for societal challenges (Kuhlmann & Rip, 2018). A transitional approach has also recently been adopted in climate policy, and an instrument such as pricing emissions of hazardous substances falls within a set of instruments and measures that aim to achieve the specified climate objectives. This emergence of new policy goals is accompanied by a debate on the types of failure that legitimise government intervention and how we view these types of failure. Box 3.1 illustrates this debate on the basis of the innovation context.

Box 3.1. Different types of failure in the context of innovation

From an economic perspective, market failure in innovation processes is the result of imperfections such as coordination problems due to transaction costs (businesses cannot identify useful innovations or customers), information asymmetry (finance providers and capital seekers do not have the same information on which to base the transaction), market power (existing businesses can hold back innovative newcomers) or knowledge spillovers. The latter concerns the advantages that innovative businesses create for their competitors without receiving compensation. Innovative businesses thus generate positive externalities, since their own investments also benefit other businesses. Innovation is also risky, so it is difficult particularly for smaller businesses to finance innovation projects externally in a situation of information asymmetry with capital providers. Against this background, generic innovation subsidy programmes have for many years made up the core of the innovation policy in the Netherlands and other OECD countries.

System failure can be understood as the consequence of coordination problems, namely problems arising in complementary goods (Rodrik, 2008). This problem is known as a coordination game in which investments generate higher returns (positive externalities) if all other relevant players make the right complementary investment. Such coordination failure can arise in various phases of the innovation process. In the early phase, when new knowledge necessary for innovation is generated, system failure occurs because knowledge is generated collectively by multiple parties: businesses, universities and semi-public institutions. If the right circumstances and incentives for this type of cooperation and knowledge sharing are not in place, there is systemic failure. The cumulative nature of the innovation process, in which future innovations build on current innovations, is also a factor here. Patents fail to achieve this 'intertemporal coordination' (Scotchmer, 1991).

In a later phase of the innovation process, the effective use of innovations depends among other things on the right investments in physical or digital infrastructure and in new knowledge and skills to exploit the innovations. Conversely, such infrastructure investments will only be made if it can reasonably be expected that they will also be used sufficiently. In addition, (possibly radical) innovations must be developed in accordance with laws, rules, zoning plans and societal views, for example by developing technical standards, user protocols and investments in human capital. System failure occurs when such coordination processes go wrong.

Transition failure occurs when a transition to a socially desirable new equilibrium does not take place or takes place too slowly. The underlying cause is often a coordination failure, as a result of which transition costs (e.g. investments) are too high for individual parties, or government failure, if the government does not intervene decisively enough to get the parties moving. As in the traditional economic view of innovation policy, transition policies can be seen as producing spillovers, although these include other spillovers (e.g. 'coordination externalities' and 'adoption externalities') rather than just knowledge spillovers (Janssen, 2022).

In transition policies, the policy goals are derived from societal challenges for which the solutions usually require new public goods (a public good in this context is a government service, but can also be a system of rules or incentives ensuring that external effects are internalised) and economic incentives or structures, whereas system policies often focus on economic competitiveness in general. The improvement of specific public goods pursued in the transition policy is politically articulated as, for example, a cleaner environment, combating climate change, public health, cybersecurity or effective disaster management. This does not alter the fact that solutions to societal challenges can also lead to new product innovations and revenue models.

In addition, in transition policy, the policy goal can only be achieved through a lengthy process characterised by uncertainty about the desired solutions, which increases the seriousness of coordination problems, both between policymakers and between the actors in the transition process in general. This requires a long-term policy perspective and has consequences for the way in which policy can be evaluated in the shorter term. The need for coordination starts early, because stakeholders must commit to concrete societal objectives in advance. This is not a trivial process, because vested interests can be compromised in transition processes. After all, transitions involve not only the construction of new systems but also the breakdown of existing systems (Loorbach et al. 2017).

In the scientific literature, a distinction is drawn between market failure, system failure and transition failure. Weber & Rohracher (2012) introduce this classification in the context of innovation policy, building on previous literature. The market perspective analyses failure from the perspective of an efficient market. The system perspective analyses failure from the perspective of a well-functioning system of businesses, knowledge institutions, governments and other parties in which the operation of institutions, routines and conduct must fit in with the desired operation and results. The transition perspective analyses failure on the basis of the transition to a new socially desirable equilibrium. In

this regard it is important that institutions, routines and conduct are updated, adapt or disappear in order to achieve the new equilibrium.

The three types of failure are generally presented as distinct failures, mainly because they have different underlying scientific traditions (Weber & Rohracher, 2012).[7] In practice, however, system and transition failures can be understood and interpreted on the basis of underlying market failure. In that sense the three types of failure do not fundamentally differ from each other. The types of failure overlap in a number of respects (coordination problems occur implicitly or explicitly everywhere), while the frameworks are also complementary: when there appears to be system or transition failure, it is often useful to ask which activities the market and existing policy are already covering, versus what remains uncovered (for instance because there is no market for it).

Market failure

For the evaluation of simple (often financial) instruments, there is an international standard and quantitative methods are in place. Government intervention is legitimised particularly by market failure.[8] The primary policy objective is to tackle underinvestment in public goods and services. The main forms of market failure are:

- *Market power.* Businesses that collude to restrict competition impede the entry of new businesses. Conversely, too much fragmentation also leads to inadequate innovation because businesses are too small to invest in new goods and services. This means the level of innovation in goods and services is too low. Sources of market power are economies of scale and lock-in effects;
- *Information asymmetry* (erroneous selection) Uncertainty about the success of investments and the time horizon within which a return can be expected makes private (or public) operators reluctant to invest. This leads to suboptimal societal investments in high-risk projects such as research and development;
- *Free-rider behaviour.* Research and development and the generation of new ideas have characteristics of a public good. There is no rivalry, which means that the private benefits for the investor are lower than the public benefits of disseminating knowledge. This means the societal investment in research and development is suboptimal due to free-rider behaviour;
- *Externalities.* Innovation investments by businesses have societal benefits for other businesses and the rest of society, although the business is not compensated for those external effects. The innovative business does not take these positive external effects into account. From a societal point of view, the business will therefore innovate too little, resulting in too few rent spillovers for consumers and knowledge spillovers for businesses and society;
- *Transaction costs.* Some markets may be inaccessible, so financing will not be forthcoming, or the search costs between parties requiring and offering financing may be so high that socially profitable transactions fail to materialise. Transaction costs may lead to coordination failure because market participants cannot coordinate complementary activities. Finally, it is possible that transaction costs will be so high that no social return can be achieved with government intervention.

[7] Government failure also plays an important role in the design of policy. There may be policy capture by lobbying or interest groups or as a result of an information deficiency leading to erroneous policy choices. We do not discuss government failure as a separate category here because it is not a justification for government intervention but a factor that mainly influences the design of that intervention, for example in the case of requirements for governance structures and transparency of decision-making that strengthens independence, or requirements for *ex ante* substantiation of policy choices.

[8] Institutional failure is sometimes added alongside market failure. This is not a justification, however, but the result of government intervention.

System failure

There is no international standard for system evaluations and the development of methods is less advanced. In addition, much less attention is devoted to causality in the sense that there are usually no quantitative methods to measure the impact of policies. The scientific literature is more descriptive in nature and mainly describes the operation of systems. Systems can be adversely impacted by restrictions. Government intervention can therefore be legitimised by system failure, which in turn may be due to market failure or government failure (Woolthuis et al., 20015; Wieczorek & Hekkert, 2012). The main forms of system failure are:

- *Lack of infrastructure.* This refers to infrastructure that is insufficiently developed due to the large size of the required investment, the long payback time and the low private return on the investment. Examples include knowledge or energy infrastructure. This system failure may also be the result of market failures like *information asymmetry* and *externalities.*
- *Capacity failure.* This may be a factor if there is insufficient capacity, making it impossible to launch new developments or activities and the system actors fail to adapt sufficiently to new circumstances. Capacity failure arises, for example, if there is insufficient access to knowledge due to a lack of or mismatch between knowledge and skills, but it can also refer to a lack of physical infrastructure. *Transaction costs* (such as search costs) and *information asymmetry* (lack of clarity as to the revenue that investment in knowledge will deliver) are often a cause of this;
- *Network failure.* Network failure can arise if network effects and network externalities are important. Network effects can lead to *market power* and cause the network to move in the wrong direction relative to societal objectives as a result of *technological lock-in.* Existing networks then impede new entrants, or networks become dependent on a dominant partner. Network failure also occurs if networks do not develop sufficiently as a result of excessively high transaction or coordination costs. Positive *externalities* from innovation then remain underexploited.

Much of the innovation policy in the Netherlands can be seen as a form of system policy. This involves instruments focused on creating an environment that is conducive to innovation opportunities, the pursuit of cooperation, the exploitation of positive externalities and adaptation to a changing market/environment. In the context of innovation, system policy focuses on the innovation activities of businesses, universities and government as the main actors in economic and innovation processes (Lundvall, 1992; Nelson, 1993).

Transition failure

Just as systems can fail, it is also possible that a transition will fail to materialise or will occur too late or stagnate. Government intervention is legitimised by the fact that the desired transitions do not materialise or do not do so quickly enough through the market, or if the market moves in the wrong direction from a societal perspective, and because the existing government policy resists changes. Forms of transition failure are (Weber & Rohracher, 2012):

- *A lack of development direction.* In the case of transitions it is not only a question of addressing the failure of the market or the system, but also of controlling the direction of change and timeliness by means of policy. This requires all actors to move coherently towards that goal or to adopt a shared vision of the transition. Symptoms of the lack of direction include, for example, a lack of shared vision about the purpose and direction of the transition; a lack of coordination between the actors or transition paths that are supposed to move the system in a new direction; a lack of laws and regulations to steer the transition and maintain its direction; a lack of funding to enable research and pilot projects; being locked into outdated routines, institutions and behaviours and a lack of infrastructure to establish routes for development paths;

- *Lack of clarity concerning demand.* The system perspective takes no account of market potential in terms of new products and services. A characteristic of transitions is that there is often a simultaneous requirement for societal, organisational and institutional changes. Failures to ensure acceptance of innovations by users and consumers are therefore a type of failure justifying policy intervention. Symptoms of a lack of demand include, for example, insufficient scope to analyse user needs and increase acceptance of innovations due to a lack of alignment with users; a lack of stimulating signals from public demand due to the government not being a user and a lack of competences to link innovations to the demand and needs of users and end-users;
- *Failure of policy coordination.* The interaction of policy fields relevant to transitions benefits from coordination, particularly policy coordination. Although the term coordination failure has been used in research and innovation policy as an example of system failure, it only refers to coordination problems of innovation actors, not to coordination problems in policy. Organising activities at national, regional and sectoral level and between different parties, for example, is important during transitions;
- *Lack of reflexivity.* This concerns continuous reflection (e.g. on strategies and assumptions) and policy assessment, as a basis for consequent revision of that policy. In transitions it is important to experiment and to learn from failures and to build on successes.

Transition policy focuses on societal objectives that must be achieved through system change. In the case of transitions there is a wish to initiate change processes, even if it is not clear at the outset which interventions work best. There is no framework for the evaluation of transition policy in the scientific literature.

The theoretical and empirical determination of the problem to which policy is the answer, i.e. identifying and analysing the market failure, system failure or transition failure, preferably combined with the development of a policy theory, is an essential step in almost all forms of policy evaluation. If it is not clear which problem needs to be solved, and why the government can possibly make a positive difference with the proposed policy, the question at the outset is whether the policy can have a positive impact – see the report of the Effect Measurement Expert Working Group (2012) for more detailed consideration of this. Although it is more common for system and particularly transition policies to be experimental, simply because there is little knowledge of which policy might work (and the societal objective still justifies the intervention), here too it is important to develop a theory in advance with regard to the operation and expected outcomes of the chosen policy. In addition, there should at least be structures (for example a formal learning cycle) that gradually provide a clearer picture of the problems that impede the achievement of the desired goals. The functioning of those structures can also be examined in an evaluation.

3.3. Effectiveness of single instruments and S/T policy instruments

Although the concept of effectiveness is unambiguous, the basis for its definition in S/T policies differs in some respects from its definition in the evaluation of single instruments.

A key difference between S/T policies and single policies is the type of action that the government undertakes. Instead of focusing on characteristics of individual actors in the system (how much does a business invest in R&D?), the actions focus on characteristics of the system as a whole (how do investments by different businesses link up?). This means policy actions consist not so much of instruments that seek to influence the behaviour or choices of individual actors, but of several instruments that seek to influence the coherence of behaviour or choices of groups of actors. This means the policy toolbox for S/T policies differs from that of interventions focused on individual actors. The core of S/T policies is the mobilisation and targeting of possibly existing instruments and initiatives. The policy is systemic because it focuses on system characteristics and not on isolated sub-

aspects. For example, networks are created (such as the MVI North Sea Energy Lab), agreements are concluded (such as the Energy Agreement), agendas are drawn up (such as the Digitisation Agenda), pacts are forged (such as the Technology Pact) and consultation committees are established (e.g. energy transition acceleration committees). The aim is to direct relevant economic and societal development by forming, directing and stimulating networks, or groups of interacting actors.

In addition, S/T policies are by their nature more dynamic than a single policy. A particular feature of S/T policies is that they seek to achieve a policy objective through continuous coordination of policy incentives and policy instruments. There are two reasons for this dynamic character:

- *Identification of problems.* First, the identification of the underlying problem is often unclear (the concept of system failure is not clearly developed in an economic theoretical framework, so in practice it remains a somewhat vague concept that can be used without making clear what the underlying problem really is), while it is also uncertain whether the chosen solution (stimulating consultation, coordination, cooperation or networking) contributes to solving the problem. As a result, system policy is often a continuous search and redefinition for an answer to the question 'what will work to improve the system and what will not?', combined with an ever-shifting range of solutions (which can be implemented by experimentation, for example);
- *Objective.* Second, especially in the case of transition policy, the policy objective and the problem shift because society continuously encounters new, different problems during a transition and increasingly understands the path of the transition, which requires a change in approach. It is difficult to ensure that policy anticipates expected weaknesses (unfulfilled conditions), because it is often unclear which transition paths will lead to the desired outcome. Hence there is also a need to continue adjusting the course and experimenting with policy to see what works and what does not work – and to prevent government failure.

Table 3.1. Differences in key principles in determining the effectiveness of a single policy and S/T policy.

	Single policy	S/T policy
Counterfactual	Construction of counterfactual based on parties outside the target group (treatment and control group).	Generally, a counterfactual is not explicitly developed in evaluations of S/T policy. Counterfactual construction may form the basis of theory, qualitative reflective studies from the literature and experiments.[9]
Direct goal of policy	Influencing characteristics and choices of individual actors.	Influencing collective characteristics and behaviour of groups of actors (in the system or involved in the transition).
Type of policy instrument	Actions addressing the behaviour and choices of individual actors. Examples include financial incentives, education or legal requirements.	Actions addressing sets of actors or a set of single policy instruments. Often aimed at processes and coordination of actors.
Complexity	Limited. Controlled through actions that engage directly with an individual actor.	Comprises multiple policy instruments, often in multiple policy areas, also focused on processes and coordination.
Target group	Clearly delineated.	Often large, heterogeneous groups of actors.
Type of failure	Market failure. Often external effects and spillovers.	System failure, transition failure. Often caused by underlying market failure and obsolete institutions.

Table 3.1 summarises the differences between the assessment of the effectiveness of policies aimed at single instruments and those aimed at S/T policies on the basis of five aspects. The counterfactual is usually not available or developed in S/T policies; the direct policy objective is aimed at the system

[9] Comparisons with other countries are sometimes possible, but here there are usually many factors that also influence the outcome and for which no controls are possible.

or the transitions; there are often multiple instruments or a combination of single instruments; making evaluation complex; the target group is often difficult to define; and the type of failure may turn out slightly differently.

3.4. Efficiency of single instruments and S/T policy instruments

Policy must also be assessed for efficiency. This has proved to be a difficult process, both in the evaluation of the policy of the Ministry of Economic Affairs and Climate Policy and in S/T policy. In theory there is a difference between three types of efficiency: technical efficiency, allocative efficiency and dynamic efficiency. Technical and allocative efficiency are static forms of efficiency, because they assess the efficiency of policy at a given point in time. Dynamic efficiency describes the development of efficiency over time. In practice, policy evaluations often only report on technical efficiency, while from a societal perspective allocative efficiency (the marginal value of an additional euro when choices have to be made between alternatives) is just as important. Dynamic efficiency is important in transition policy, because new information becomes available over time about the best way to achieve the policy goal (and which policy should be associated with it).

Technical efficiency

Combating waste is a means of increasing technical efficiency. Waste of public resources occurs when the effort is greater than that needed to achieve a goal. Preventing waste means the effect of policy remains the same, but expenditure decreases. Other examples of improvements in technical efficiency are preventing errors and streamlining processes.

Reducing expenditure is not the only way to increase technical efficiency. Technical efficiency can also improve by delivering higher quality for the same cost, for example because new technologies are used or if more can be achieved with the same resources. Policy that is accompanied by lower costs is therefore not by definition equivalent to technically more efficient policy (after all, it may also be at the expense of quality) and an improvement in technical efficiency is not by definition the same as lower costs (an investment, for example in a new ICT system, yields a higher quality).

In evaluations, technical efficiency revolves around the relationship between the benefit (effects) and the cost of policy. The cost measures the value of the public resources used and the benefit measures the value of the effect achieved. Technical efficiency among alternative policy options can be compared by examining the net benefit per spent euro, with a higher benefit per spent euro equating to higher technical efficiency. In practice, analyses of technical efficiency often focus on cost: can the same services be provided at lower cost?

Allocative efficiency

Allocative efficiency can be improved by deploying the available public resources where they deliver most for society. Allocative efficiency applies within system and transition policies and to the total expenditure of the system or transition.

The allocative efficiency of the system itself can be improved by deploying public resources where they generate the biggest effect within a system. This can be done by implementing policy with the highest technical efficiency or by deploying policy within a system with the highest efficiency of implementation (this also includes the level at which policy is made and implemented: national or regional, central or decentralised and top-down or bottom-up).

Allocative efficiency also plays a role at the level of the system as a whole. Suppose, for example, that an additional euro generates more prosperity if it is spent on the digitisation agenda than on climate policy. In this example, it is allocatively efficient to invest in the digitisation agenda, even if from a

technical perspective (positive net benefit) the euro could also be spent efficiently on climate policy. The investment in the digitisation agenda leads in this example to higher common benefits than an investment in climate policy. Determining the level of policy efforts is therefore an allocative matter, in which the benefits generated by the system must be assessed against the benefits of other public goods or societal objectives.

Allocative efficiency of system and transition policies can be evaluated on the basis of two questions:

- Which system and transition policy is allocatively efficient? In order to answer this question, information is required on the benefits and costs of all the S/T policies. Benefits are often difficult to measure, but often it is also impossible to allocate costs clearly to a specific policy. For example, coordination programmes often use existing instruments, so allocation of costs is not immediately clear. In practice it will be difficult to answer this question, since the respective benefits of different policy initiatives have to be assessed relative to each other.
- What level of public spending is allocatively efficient? To answer this question, it must be borne in mind that efficiency may vary with policy intensity, specifically in the case of the coordination efforts associated with S/T policies. Too little commitment to coordination will not be effective and the effort will be in vain, but too much coordination will result in only insiders being involved.

Methods that can be used to measure allocative efficiency are available to a limited extent, because a comparison must be made with the return on resources for alternative purposes. In the current practice, studying case studies is actually the only suitable method, because the other methods focus on one area of application. A risk here is that all reallocations are assessed as positive, whereas it is not clear whether they will indeed be effective.

Dynamic efficiency

Both technical and allocative efficiency are static dimensions of policy, as they evaluate effectiveness at a specific point in time (usually the end of a policy period or a mandatory evaluation time). Technological progress will make it possible to achieve greater efficiency in the future than is now possible with the current knowledge and policies. Dynamic efficiency is important for evaluating system and particularly transition policies, because it describes how current policies will affect efficiency in the future. In the future, more room for innovation could lead to new solutions for climate problems, for example, that save costs or help to meet climate objectives more efficiently.

Stimulating change plays an important role in system and transition policy. Innovation is aimed at new products and services and types of coordination, production processes, communication or organisation that make the system work more efficiently. This may include technological innovations, such as digitisation, organisational innovation, such as logistics processes, and external organisational innovation, such as outsourcing of production. The goals of S/T policies consist of improving processes, which can involve primary (products and services) and supporting (governance) processes, and achieving societal objectives more efficiently.

The determination of efficiency takes future benefits into account and compares these with the costs. The cost-benefit ratio may change over time. Policy at the start of a transition costs a lot of time and resources, while the benefits lie far in the future. Benefits become visible when the transition shows an exponential development trajectory (for example in performance indicators such as savings achieved in terms of CO_2 emissions). The efficiency of the policy consequently varies at different points in time, particularly if the coordination policy can be scaled back after a tipping point has been reached.

A complication in the evaluation of systems and transitions is that technical efficiency (lower cost with the same benefits or higher benefits for the same cost) is not necessarily consistent with dynamic

efficiency. For example, there may be a technological lock-in and path dependency if, on the basis of technical efficiency, a choice is made that provides little scope for new technological developments that are ultimately more efficient (Foray, 1997).

Indications for efficient policy

When evaluating the efficiency of S/T policies, it is important to take account of the dynamic nature of the policy. The evaluation of dynamic efficiency is complicated by uncertainty, however. On the basis of option theory, there are at least three possible ways to measure dynamically efficient policy: (1) incorporating the option to wait before introducing policy instruments, (2) dividing choices into a number of steps to allow interim adjustment (or discontinuation) of the policy, and (3) experimenting on a small scale with new opportunities to learn which policy is the most effective and efficient. All these aspects of policy relate to dealing with uncertainty about the policy to be pursued, which is an important feature of S/T policies. In order to arrive at a judgement on the efficiency of certain policies, it is therefore possible to examine the extent to which these elements actually form part of the policy, such as subsidy schemes that are only used when market developments give cause to do so, interim evaluations followed by adjustments or policy experiments actually implemented.

Waiting

If a technology or solution to an issue is available, it does not have to be deployed immediately as a policy instrument. Uncertainty about the efficient use of the technology is caused by uncertainty about the price of different technologies to resolve the issue. A higher degree of uncertainty leads to the postponement of choices.

Policy efficiency can be determined in an evaluation by assessing the extent to which information has been gathered to determine the current value of a technology and weighing it against possibly more efficient technology in the future. An example is the current effectiveness of solar panels to achieve a climate objective. This value fluctuates due to changes in the market for solar panels and the availability of alternatives to solar. In the case of solar panels, for example, it is about the price of raw materials, the constantly improving technology, the number of manufacturers and the possibilities of using alternatives, such as wind energy. Development costs are also incurred when investing in a particular solution or technology (in option terms, this is the exercise price). These costs differ depending on the technology and the efficiency evaluation should identify the extent to which the use of resources has been assessed. The resulting assessment must also include a time dimension, because the achievement of objectives must not take too long. If a more efficient technology is not expected to become available in the foreseeable future, it is possible to evaluate the extent to which this decision has been made on the basis of sufficient available information and what costs and benefits will result.

Division of investments

Assessing the sequential investment process is a second aspect of policy that can be examined in the context of dynamic efficiency. Dividing up investments in technologies with an uncertain return is something that happens in many sectors, such as aviation and pharmaceuticals. In transition policy, the division of the process leads to decision points. A range of technologies are available to achieve the climate objectives. It will probably be the mix in the development and use of these technologies that determines the success of the policy. In an initial phase of assessing efficiency, choices must be made among a number of promising technologies, which can be tested or experimented with in a second phase. The resulting choices can be assessed for technical efficiency with regard to the policy cost, for allocative efficiency with regard to the option of halting or intensifying the policy and for dynamic efficiency by evaluating the extent to which uncertainty has been explicitly factored into the policy assessment. A second phase concerns the building of a structure in which cooperation takes place between public and private operators or public investments are made. Technical efficiency

assesses the extent to which a waste of resources has been prevented and existing structures and instruments are used where possible. Subsequently, the implementation of policy can be evaluated for technical efficiency (preventing waste and improving quality) and for the options of halting the policy (if it is not sufficiently successful) and adjusting it (if new information is received or if interim goals have been achieved).

Experimentation

A third way of achieving efficient policy is to conduct (possibly small-scale) experiments to find out which policy is effective and efficient. Policies that use such methods are more likely to be dynamically efficient than policies involving no experimentation. It is important, however, that the experiments themselves are also set up effectively and efficiently, and that lessons can actually be learned from the experiments conducted at the level of the system or the transition. It must then be clear that the results are used to shape policy.

3.5. Challenges for the evaluation of S/T policies

The purpose of policy evaluation is to determine whether the policy at a particular time is effective and efficient, so that the next step can be taken (continue, adjust or stop). An important step in evaluating effectiveness is formulating testable hypotheses to ascertain the effectiveness of the policy.

Such hypotheses are based on an action (e.g. a subsidy, creating a network, starting a coordination meeting, laying down a joint vision for the future), an effect which that action aims to achieve (more innovation, better coordinated research, a shift towards innovative projects) and a mechanism through which the action leads to the desired effect (the cost of investing in R&D is reduced so that more innovation takes place, better coordination leads to innovations that reinforce each other).

One of the challenges in evaluating S/T policies is separating the effects of the individual elements from the effect of S/T policies as a whole. S/T policies often also contain several standard policy elements that focus on the behaviour of individual actors, such as subsidy instruments. In practice, the effects of S/T policy and the effects of single policy instruments are constantly confused. What happens in practice is that S/T policies seek to move a number of policy instruments more in the same direction. Take the digitisation agenda, for example. The effectiveness of S/T policies in relation to this agenda is determined, for example, by examining outcome measurements related to the degree of digitisation, such as the number of broadband connections or the degree of digital literacy. However, the S/T policies in the field of digitisation consist of a collection of 'ordinary' and, *inter alia,* coordinating activities and incentives for parties to work together. In order to determine the effect of S/T policies on digitisation, it is necessary to identify the added value of the S/T elements relative to the traditional elements. How much do those additional activities yield, compared to a world in which they would not take place? Which functions in the operation of the system need to change? And what can the mix of instruments in the field of digitisation do about that? Box 3.2 illustrates these challenges on the basis of the evaluation of the Top Sector approach.

This is the crux of the challenge for the evaluation of S/T policies. Evaluating precisely what that better coordination will yield for a better functioning system compared to a world in which that coordination would not take place is far more complex and complicated than evaluating the effect of a single instrument focused more or less directly on influencing the behaviour of individual entities.

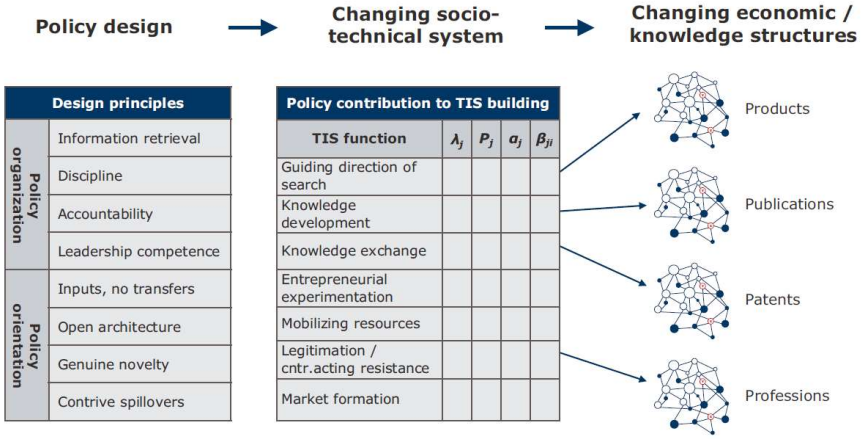
The essential point is that a combination of interventions should promote conditions that lead to system improvements or desirable transitions. The extent to which policy makes an effective and efficient contribution to this requires an understanding of the interplay between all mobilised or adjusted interventions, as well as the various system changes they bring about. Due to the multiple

interventions, the relatively high importance of governance structures and processes (compared to instruments or a budget article) and the long effects chain, the scope of evaluations is therefore broad.

Stern et al. (2012) find that the lack of appropriate methods for system policy evaluation can be traced back to the considerable complexity of the set of actions and interventions involved. Determining causal effects is made difficult by the large scale, broad scope, nested structure, multidimensionality and multi-level character of system policy. In addition, the assessment criterion changes. It is not a question of realising a goal, but of functional changes in the system that determine the extent to which the operation of the system improves or the transition moves in the right direction. In this context, S/T policies are judged more on the basis of the resulting movement than on the basis of the level attained at a certain point in time.

Box 3.2. Evaluation of the Top Sector approach illustrates challenges

The evaluation of the Top Sector approach illustrates the challenges involved in evaluating S/T policies (Janssen, 2019):



The evaluation of the governance structure took the form of identifying and then assessing policy principles that were deemed important for this specific policy approach. An analysis of the literature highlighted eight policy principles that are deemed relevant to the creation of conditions whereby governments, businesses and knowledge institutions can explore and strengthen new, promising innovation paths. However, these policy principles were not based on empirical research showing that innovation outcomes would improve if this principle were fulfilled.

The policy evaluation also included an assessment of the connection between the initiatives for system strengthening. The key question was whether the initiatives were consistent with the main issues in the systems on which the Top Sectors focused. The analysis was conducted on the basis of a framework of key processes of importance for a well-functioning technological innovation system. The research also included an attempt to make statements on how well the interventions worked. In view of the multiplicity of interventions for each of the nine Top Sectors, it was not possible to conduct an empirical analysis of these.

It was not clear what an effect measurement of the Top Sectors approach should relate to. The ultimate policy objectives consisted of increasing the public-private expenditure on PPP projects, increasing the general R&D intensity in the Netherlands and raising the Netherlands in the ranking of competitive countries. The first policy objective is most closely connected with what was actually undertaken in policy terms. At the time of the evaluation it was not possible to say anything about the actual outcomes generated, partly because it was impossible to relate the recent introduction of the system approach to macro outcomes such as higher R&D intensity.

Generally, carrying out pure effect measurements requires visibility on which intervention was or was not applicable to whom, and how their performances have developed over time in terms of level and direction (preferably in relation to the pre-intervention situation). In the case of S/T policies, the system as a whole generally does not fulfil these conditions (although effect measurement may be possible for subsystems or parts of the system policy), for example because:

- The precise target group to which policies apply is often not clearly defined. It is not always clear who 'obtains' something from the policy or is influenced to a certain extent. The fact that the

approach relies on a policy mix is not necessarily a problem, as long as it is known for each instrument who has received which (positive or negative) incentive. However, policies that increase the conditions for innovation do not (always) act on organisations; for example, they may act on regulations, infrastructure or knowledge intended to contribute to innovations for a certain transition. Effect measurement requires a picture of differences in the extent to which individual or groups of organisations are affected, because that differentiation forms the basis for explaining performance differences. If policies comprise multiple instruments focused on system factors (instead of support for specific target groups), it can become very difficult to determine who has benefited or suffered from the policy and to what extent, and who can therefore form a good experimental or control group;

- As has been established, finding a counterfactual is often difficult or even impossible, because there is usually only one system/transition in each field. In interventions at organisational level, it is usually easier to identify treated and untreated organisations that can be compared over the same period. For measurements at the level of ultimate objectives such as bringing about a circular manufacturing industry, there are likely to be only a few rough comparisons over time or with other countries; these are ill-suited for making statements about the specific influence of policy;
- There is not always agreement on precisely what policy instrument to evaluate when considering the evaluation of system policies. System policy often consists of a combination of various instruments in different policy areas, with instruments also being very diverse in nature. System policy then focuses on coordination or coherence between the instruments to improve the operation of the system. Evaluation questions relating to system policies should therefore be clearly formulated in terms of the instruments to be evaluated in conjunction with each other and to what extent the impact of policy on coherence or coordination itself forms part of the evaluation. Even if there is agreement on concrete interventions that are important for system strengthening or a transition, it may still be difficult to identify which aspect/part of this can be linked to the S/T policy. Even the budget involved in certain system policies can therefore be unclear, which make it difficult to make statements about the efficient use of resources;
- Finally, S/T policy is sometimes aimed at boosting developments that are not yet taking place or at making transitions characterised by non-linear effects, in which the effect of policy can fluctuate greatly over time. Both aspects make it hard to say whether an outcome is already sufficiently in progress.

4. Evaluation framework for system and transition policy programmes

This chapter presents a comprehensive evaluation framework for S/T policies. We first introduce principles to clarify what the key assumptions and principles are when evaluating S/T policy. We then present step-by-step perspectives and frameworks from the scientific literature that can be useful for evaluation.

4.1. General principles

There are a number of key principles with regard to evaluation.

Policy theory. A failure must have occurred. Based on the failure, a clear description must be drawn up of the policy instrument to be used, the policy objective pursued with the policy instrument must be defined and the mechanisms by which policy efforts should lead to the desired policy objective must be determined.[10] It is relevant here to analyse the spectrum of policy efforts and instruments, and to examine for each of them which part of the policy theory is associated with it when it comes to the *input – throughput – output – outcome – impact* chain. It must also be clear what the financial consequences are for the government (and if relevant for other actors) and how government failures are prevented. This makes it possible to switch between different observation levels; effects lower on the axis of validity of policy attribution (causality) are often easier to assess than higher levels. In this first step it is important to analyse to what extent the policy effect is only the sum of the effects of individual policy efforts or whether there is synergy.

A crucial ingredient of policy theory is to identify what is needed to draw conclusions about the effectiveness and efficiency of S/T policies.[11] We distinguish the following levels on which the evidence for hypotheses can be tested: consistency, falsification, correlation, and causality. There is consistency if observations are consistent with the assumed mechanism (fewer diesel engines are sold). There is falsification if it is possible to find counterevidence for the policy theory (sale of diesel engines has risen despite the policy). Correlation exists if observations of the policy instrument used coincide with desired outcomes (a decrease in CO_2 emissions coincides with a decrease in the sale of diesel engines). Causality exists when observations are the cause of desired outcomes (the pricing of CO_2 emissions has made diesel engines less attractive and sales have fallen). This four-part assessment can be made for each level (input – throughput – output – outcome – impact), with the aim of establishing causality. This is shown in Table 4.1.

There is a hierarchy here in which causality ranks above correlation, correlation above falsification and falsification above consistency. In the case of single policy instruments, in which the policy objective is clearly defined, the target population is known, the counterfactual can be identified and the instrument is not continuously adjusted, there are in principle good preconditions for determining causality. After all, there is a visible counterfactual, such as a control group of organisations that resembles the treated group as far as possible and whose performance trajectory is also known. Sufficient data of appropriate quality are also available to evaluate the instrument for effectiveness.[12]

[10] This should be done not only for economic reasons, but also to comply with the Dutch Central Government Accountability Act. Section 3 of that Act states that proposals, intentions and commitments must include an explanation regarding, among other things, the objectives, effectiveness and efficiency that are being pursued.

[11] The Maryland Scientific Methods Scale is well known. In the context of evaluations, it is also referred to as the 'effect ladder'. The ladder metaphor indicates that there is a hierarchy in the appropriateness of certain research methods for making statements on policy effects.

[12] The practice here falls short of the theory. Policy that could in principle be evaluated in accordance with this system is often not evaluated. This is because such evaluation requires great effort (and hence such evaluations are more costly) and there are high risks that the evaluation will deliver unwelcome outcomes. These evaluations are also often somewhat one-dimensional and limited in scope, so policymakers do not feel they have been given tools to improve the policy.

Table 4.1. Effects matrix to determine which step on the effect ladder is possible for each identified mechanism

	Input	Throughput	Output	Outcome	Impact
1. Consistency					
2. Falsification					
3. Correlation					
4. Causality					

Data. What data are available for a policy evaluation determines which aspects of policy can be measured. System policy consists of instruments and often also addresses coordination problems and coordination activities. The existence of structures with well-functioning governance is an example of this. Another example is the extent to which policy adjustments result in intensified or revised coordination. This makes it possible to identify on the basis of the effects matrix the level on which an evaluation can focus and which methods can be used.

Type of evaluation. Evaluations of S/T policy do not only require a standard effect measurement, since the policy derives its legitimacy from the outcome it generates, but mainly focuses on a combination of instruments (which should reinforce and/or complement each other) and on directing the processes that lead to this outcome. There are two questions here. The first is to what extent interventions lead to the desired adjustment of processes and contribute to the desired transition path. The second is to what extent the adjustment of the processes leads to the desired outcome. Both are also forms of effect measurement, but on other outcome variables. The problem in the case of evaluations of S/T policies often lies in the second question, because that question cannot usually be answered causally with the available methods.

The evaluation of S/T policies is to some extent focused on input and process, because the government is trying to tackle a societal problem. A precondition for effectiveness is then the establishment by the government of processes that make it plausible that a relevant contribution will be made to the resolution of the problem (Boon & Edler, 2018). A judgement on the established processes thus constitutes a step in the evaluation of S/T policies. In practice, this plausibility is estimated by using theoretical and qualitative-analytical methods to outline the conditions that effective policy should meet. In that case, evaluation is an exercise in verifying whether the policy actually meets these conditions. The limitation of this approach is that it merely shifts the problem of measuring the effects of policy. It is important to demonstrate that the conditions described as effective in the literature are also effective when applied to the specific Dutch context (external validity). The scientific literature on S/T policy effectiveness is largely non-empirical or based on fragmented case histories. It therefore scores low for causality in Table 4.1.

Hence there are increasing calls for formative evaluation (Molas-Gallart et al., 2021), possibly combined with a summative evaluation for accountability (Magro & Wilson, 2019). The concept of formative evaluation has its origins in education. Formative evaluation focuses on progress towards the final objective and feedback on the learning process, while summative evaluation focuses on whether or not the final objective has been achieved (or sufficiently achieved). Due to changing circumstances, it is often insufficient for S/T policies to judge whether policies have been effective in the past, as effectiveness at a particular point in time does not guarantee that policies will continue to be effective when new system and transition issues arise. An example is the observation that the Top Sectors approach has helped boost new collaborative relationships, but that continuing to build these generates diminishing returns that make only a limited contribution to achieving the intended policy objectives. When even more new cooperation structures have been created, it will be more effective to tackle other system factors. It is therefore particularly important to know how the system is faring, which policy fits in with it and whether the organisation and processes are functioning effectively. Previous policy experiences and scientific literature provide tools for assessing this

effectiveness. This perspective is consistent with the idea that methods scoring highly for causal evidential value can be supplemented with other evaluation methods.[13]

A characteristic common to some of the alternative evaluation methods is that there is more room for involving policy implementers in understanding the intended and actual functioning of policy, and the factors that influence it (Barbrook-Johnson et al., 2021). For example, policy implementers can participate in defining and applying criteria for process evaluations, although it is not certain that this is desirable from an independent evaluation perspective. The joint design and implementation of evaluation research fits in with the approaches of developmental evaluation (Patton, 2011) and reflexive evaluation (Verwoerd et al., 2020). Van der Knaap et al. (2020) talk of responsive evaluation (where there is a constant need for direction and interpretation in the interests of various stakeholders) and of contextual-realistic policy evaluation (focusing on the question of by what mechanisms and under what circumstances policies work). Arnold et al. (2018) consider the tools that such evaluation perspectives offer for the evaluation of complex innovation and transition policy approaches. They point out that advancing insight plays an important role in system policy, because it is not clear in advance which intermediate goals and policy efforts will deliver the ultimate policy objective. Monitoring progress and achieving goals is therefore very important for assessing effectiveness as well as for determining (dynamic and allocative) efficiency.

4.2. Policy theory

In this section we discuss a generic policy theory that should serve as a starting point for any evaluation of system and transition policy (Figure 4.1, based on Janssen et al., 2022).

For actual applications, this will of course have to be refined to provide sufficient points of reference for evaluations. Before an evaluation takes place, it is a necessary precondition that a detailed policy theory has been drawn up. A policy theory should preferably be developed before policymakers embark on their S/T policies. Without a predefined policy theory, there is a risk that an evaluation will ultimately be no more than a description or attempt to legitimise what has taken place (also because insufficient data will probably be available). The policy theory for S/T policies consists of a number of elements.

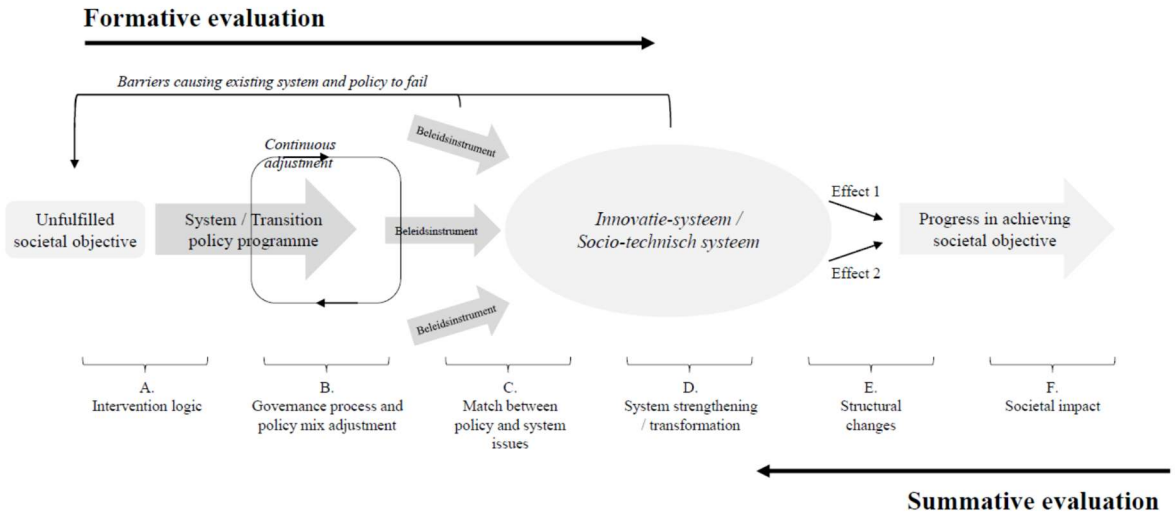
- *Societal objective to be achieved.* The starting point of the policy theory is a societal wish or requirement that cannot be fulfilled under the existing state of affairs (e.g. sustainable energy generation). As a starting point, it is important to determine the relevant failure that explains why the societal objective will not be achieved without government failure (for example, no or insufficient pricing of environmental pollution leads to a lack of sustainable energy generation, resulting in excessive environmental pollution).
- *Interventions.* S/T policies seek to change this through a comprehensive programme that aims to bring about change by means of synergistic policy instruments (e.g. the energy transition).
- *Coordination of efforts by means of the S/T policy programme.* The coordination relates to new and existing new policy initiatives that affect, change and optimise a system. The policy initiatives may be specific to a certain system factor (such as knowledge development, market formation and human capital) or have a systemic scope (such as consultation and coordinating financial incentives).
- *Continuous adjustment to the policy mix based on new knowledge.* There is a continuous adjustment or feedback loop whereby the programme is adjusted in order to achieve the

[13] Van der Steen (2020). Task-oriented evaluation: methods that fit the policy issue. P. 35 in: Beleidsevaluatie in theorie en praktijk (Van der Knaap et al., 2020).

objectives. Some instruments are more effective than others and new technological and societal developments are changing the effectiveness of the interventions.

- *Measuring outcomes and interim outcomes.* S/T policies seek to bring about changes in a system by means of synergistic instruments. The strengthening or transformation of a system is already an interim outcome in itself. This must then lead to effects related to a policy objective. For example, research and market participants collaborate more often when conducting research and focus on subjects that the government has defined as relevant, or, for example, low-priced solar panels appear on the market over time that help to make energy generation more sustainable.[14]
- *Monitoring of transition paths.* On the far right of Figure 4.1 there are observations on how the transition is actually proceeding. Indicators must be drawn up to monitor the transition path and obtain a picture of progress. These indicators in themselves do not yet say anything about the influence of policy.

Figure 4.1. Generic policy theory for S/T policies, with two associated evaluation perspectives



4.3. Two evaluation perspectives

In the literature on S/T policies there are a number of studies calling for a combination of evaluation methods. Turnheim et al. (2015), for example, describe the complementary analyses that are relevant to research into different scales on which transition paths are developing (initiative-based niches, and the socio-technical system as a whole) and Arnold et al. (2018) propose evaluation approaches linked to different levels of policy responsibilities (such as ministries, executive agencies, programme managers and project leaders). A recent OECD study into mission-oriented innovation policy (Larrue, 2021, p. 14) states that, despite a clear trend towards system policy, existing evaluation instruments and techniques are primarily suitable for individual programmes and instruments. The lack of system-wide evaluation methods is mainly due to the complexity of effect measurement and attribution in policy characterised by large scale and scope, the nested structure whereby policy is part of a larger set of objectives and a multidimensional, layered design.

The limitation of the evaluation methods proposed in the system and transition literature is once again that it is not primarily aimed at the accountability question of whether policy is effective. The main

[14] A relevant point here is that there may be policy objectives on different levels, such as lower CO_2 emissions, production of more solar panels, a larger proportion of sustainable energy production or increasing costs of CO_2emissions. All these objectives can legitimately be part of the policy theory.

basis is a review of principles that are assumed to be a precondition for effective policy. When policy is assessed against these principles, the quality of that assessment depends greatly on the substantiation of those principles. An example of a principle is that the control in the case of innovation policy must be in the hands of 'open' structures involving very divergent actors (Schot & Steinmueller, 2018). Although the literature indicates that this principle helps to prevent a small group of established parties from calling the shots (network failure), there are situations in which this principle is ineffective (falsifiable). Principles are bases that are assumed to provide guidance, but they are not laws. Another example of a principle is that instruments must be consistent with each other and with the overarching policy strategies to which they belong (Rogge & Reichardt, 2016). This is a principle that is recognised as difficult to comply with in practice. After all, policy strategies often draw on existing instruments and processes that impede a new optimum policy mix. Formative evaluation is therefore not only about determining the extent to which the policy complies with relevant principles (the answer will often be negative), but also about determining where imperfections lie and how these imperfections can be resolved or reduced.

Based on a generic policy theory, formative evaluations can be said to be particularly suitable for forming a picture of the (potentially influenceable) processes, structures and mechanisms that are intended to lead to system change and transitions. Summative evaluations draw conclusions about the extent to which policy developments arise that are of importance for a policy objective.

Formative or normative evaluation

Formative evaluation focuses on (possibly interim) evaluation of progress towards an ultimate objective and thereby learning about the functioning of policy, so that adjustments can be made. In the case of aspects such as research design, data collection, interpretation and 'receptiveness' to outcomes, participative methods are often chosen (e.g. intensive involvement of policymakers and implementers). It is important, however, that the independence of the research is properly safeguarded, both in the formulation of questions and in the implementation.

Formative evaluation lends itself, among other things, to answering questions about the quality and suitability of control mechanisms and governance on which the S/T policies are based. There is usually no ready-made policy framework that indicates which process characteristics and indicators are relevant to a policy approach. First, it is necessary to find out what the design principles or standards are against which the process can be tested. Two essential steps are *identifying* principles by researching literature and/or consulting parties who have been closely involved in the design (and adjustment) of the policy; and testing the principles with more stakeholders (not only designers), preferably with a group of critical partners and stakeholders and independent experts.

The key principle for the formative evaluation is to find out which activities and structures have been set up by policymakers to fulfil the design principles. For example, if it is important that the decision-making structure leaves room for participation by parties with solutions that do not operate at the heart of the system, it is not sufficient simply to sound out those involved to find out what they think. It is crucial to know the principles and, as an evaluator, to form an independent picture by investigating which activities have been undertaken to gain a picture of this participation (with networking events, for example), to weigh that participation (for example, by setting up a neutral advisory board to ensure participation) and to open up decision-making procedures and interventions for this purpose (e.g. with instruments).

Summative or positive evaluation

Effect measurement is a powerful instrument. Experimental or quasi-experimental designs enable the purest possible effect measurement, so it may be justified to focus the analysis on only part of the scheme or instrument, of the target group, of the years in which the policy ran, of the goals/effect

types according to the policy theory etc. This will be the case, for example, if a competitive scheme uses scores and a comparison is made between applicants who receive an unsatisfactory score and applicants whose score is just adequate. The focus on effect measurement with the highest possible validity may be at the expense of a more comprehensive (but less robust) insight into what the scheme has delivered across the board. Therefore, under the heading of outcome evaluations, it may be desirable to use additional evaluation techniques to make the effect measurement more complete.

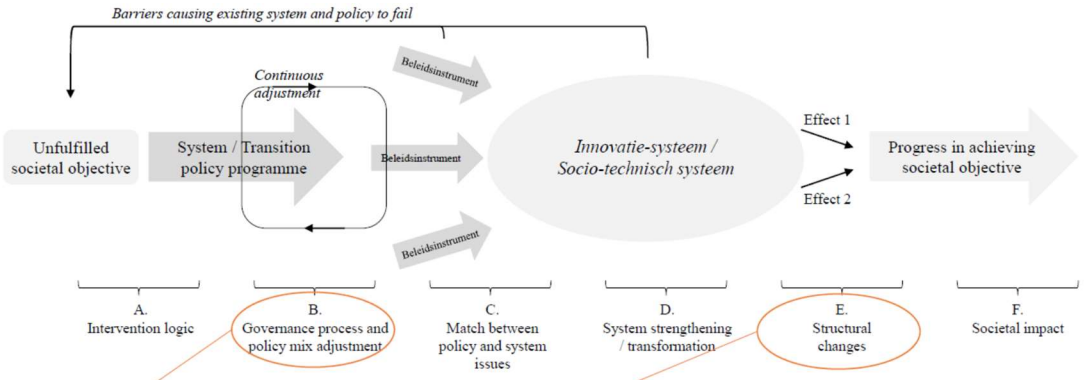
Once it is known how individually effective various schemes are, there will be partial visibility on the extent of system change involved. Ideally, an effect measurement should indicate whether the generated outcomes amount to circumstances that are favourable for the ultimate policy objective to which the policy is intended to contribute. It is sometimes possible to examine this empirically. For example, if it is known which parties or innovation directions would benefit from the policy, the question is whether they have also experienced tangible benefits from the initiated system change. The extent to which that system change can be traced back to policy efforts by means of attribution analysis is relevant, but even if it cannot be traced back, it may be useful to know at least which aspects of the system are working adequately and which are not. There is a tension when it comes to the effectiveness of policy and measuring the extent to which certain conditions exist (regardless of evidence of policy effect).

4.4. Theoretical assessment and analysis frameworks for S/T policies

There is no universal S/T evaluation framework. With a summary of the available scientific literature and evaluation practice, a number of perspectives can be distinguished that are relevant for the evaluation of S/T policy (Janssen et al., 2022). This concerns the six elements shown in Figure 4.2: (A) intervention rationale, (B) governance processes/structures and the policy mix, (C) connection between policy (mix) and issues, (D) system strengthening/transformation, (E) structural change and (F) societal impact.

For each specific perspective it is possible to draw up detailed hypotheses indicating what change can be expected with regard to that aspect of policy theory. The effect ladder can be used to determine at what level (input – actions – output – outcome – impact) conclusions can be drawn, the extent to which these conclusions are causal and what data are required to be able to draw conclusions about effectiveness and efficiency. In practice, it will not be possible to enter all horizontal or vertical goal types. It is important to show which conclusion can be drawn for which elements and how the whole balance sheet of possible conclusions on effectiveness and efficiency of S/T policy looks (Figure 4.2).

Figure 4.2. Graphic representation of the evaluation system



Hypothesis X	Input	Actions	Output
1. Consistency			
2. Falsification			X
3. Correlation	X	X	
4. Causality			

Hypothesis Y	Input	Actions	Output	Outcome	Impact
1. Consistency					
2. Falsification					X
3. Correlation	X	X			
4. Causality			X	X	

Hypothesis X: *"Coordination leads to adjustment of conflicting policy"*

Hypothesis Y1: *"The policy leads to new cross-sectoral collaboration patterns"*

Hypothesis Y2: *The policy leads to a shift of people and resources to areas where they have most impact (= thematic convergence in innovation efforts)"*

A. Intervention logic

The evaluation question that determines the intervention is whether the policy is legitimate. To determine this, different types of failure can be used as a guideline (see the discussion in Section 3.2). As with regular policy, examining intervention logic is not a matter of stating what type of failure might be involved, but rather examining evidence that a particular problem is occurring and the extent to which policies can reduce or even solve this problem. It is easy to *assume* that entrepreneurs do not use a particular sustainable technology because, for example, they do not properly recognise its benefits (market failure: information asymmetry) or because an infrastructure for it is underdeveloped or start-up costs are too high (transition failure: lack of a clear direction). Both examples of failure require a different intervention. This means that naming a potentially relevant type of failure is the starting point for demonstrating that the intervention addresses an issue that it can actually remedy or reduce.

When it comes to S/T policies, policymakers are bound by frameworks, a diversity of goals and interaction with parties having different goals. The degree of policy coordination and reflexivity will never be optimal, or at least not for long. Identifying the reasons why a market, system or transition failure occurs and why remedying it is effective requires an integrated programme rather than a targeted individual intervention. Reconstructing policy theory in this way therefore partly amounts to identifying (and assessing) the logic of the chosen policy design. From the point of view of the intervention logic, it should be obvious to combine several incentives in a single programme and these incentives should reinforce each other, requiring coordination at the system level.

B. Governance processes/structures and the policy mix

The scientific literature provides tools for assessing the quality of the established governance processes and structures and the extent to which these are effective in setting up a policy mix. Consistency and coherence are keywords for governance and structure: there must be no internal contradictions in how the policy is set up and what it aims to achieve. Judging which governance is effective in practice is an empirical question that is answered qualitatively in the scientific literature. To do this, it is often first necessary to determine which principles the design of the policy must fulfil in order to be effective at all. For example, Grillitsch et al. (2019) investigated the Swedish strategic

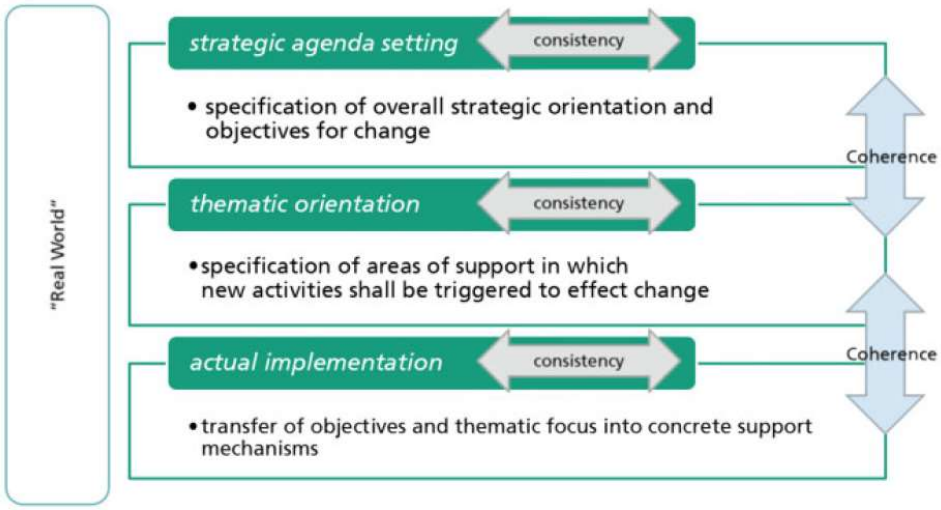
innovation programmes by examining whether the governance was in line with four transition failure criteria obtained from the transition literature (see Section 3.2).

Box 4.1 briefly discusses the general framework that can be used when evaluating the governance of most S/T policies. This amounts to checking consistency at different levels of policy development and the coherence between them. Rogge and Reichardt (2016) distinguish three levels: strategic, thematic and implementation. Not only must components at every level be free of contradictions and mutually support each other, but the policy must also be coherent between levels.

Box 4.1. Tools for process evaluation in S/T policies

A recent article by Kroll (2019) provides a structure for systematically checking the extent to which the governance of the policy implementation is proceeding effectively.

He distinguishes three levels in the translation of policy ambitions into policy implementation. Besides the fact that activities and structures must be consistent at every level (no internal contradictions), they must also be coherent between different levels. This can be jeopardised by actor-based challenges (incompetence, irrationality), but also by functional challenges such as the lack of information, resources, mandates and organisational skills.



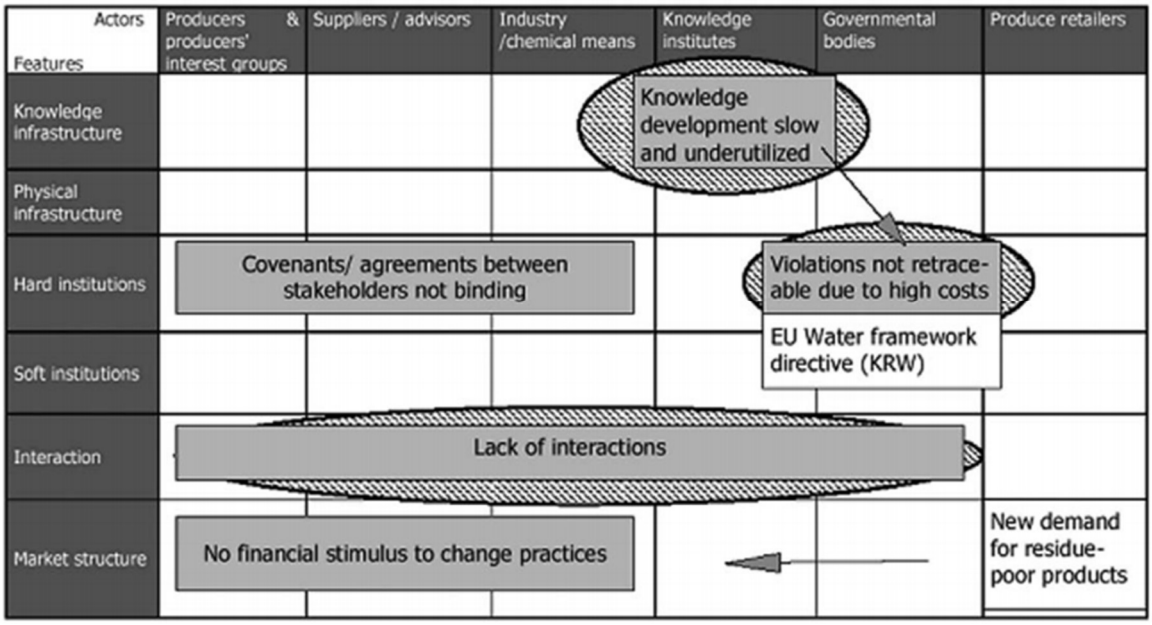
In addition to evaluating the governance processes themselves, it is also possible to evaluate how those governance processes lead to adjustments to the policy mix. Reflexivity is important in S/T policies: the ability to adjust incentives on the basis of monitoring information and on the basis of any new issues that arise. Again, there is no uniform standard for determining 'good' and 'bad' policy mix adjustments, but there are examples. The Frauenhofer Institute recently published a report in which evaluators examined whether policy instruments are consistent with the 'impact pathways' that guide the missions in the German HighTech Strategy 2025. The (tailor-made) analysis framework here is aimed at determining consistency, in this case between the revisions in instruments on the one hand and visions of how objectives can be achieved on the other.

C. Match between policy (mix) and issues

The third perspective concerns the match between the policy mix and issues in the form of weak/missing system factors or transition processes. This requires a picture of issues at the detailed level. According to this perspective, the valuation should indicate whether the governance and interventions result in incentives that, in terms of scope and dose, match the issues that impede the desired changes.

Van Mierlo et al. (2010) study system issues in greenhouse horticulture – see Figure 4.3. They identify issues by looking at actors and elements in the innovation system. In much the same way, the evaluation of the Top Sectors approach involved an examination of weak functions and key processes in the technological innovation systems corresponding to the Top Sectors and the extent to which initiated actions were in line with them (Ministry of Economic Affairs and Climate Policy, 2020). A more recent example is an investigation into barriers in the mission-oriented innovation system relating to sustainable shipping (Wesseling & Meijerhof, 2021).

Figure 4.3 Example of analysis of alignment between anticipated policy mix (the circles) and system issues (the squares). Source: Van Mierlo et al., 2010.



D. System strengthening/transformation

There is a tradition of studies investigating system strengthening by looking at structures and 'key processes' in technological innovation systems (TIS) (Hekkert et al., 2007; Bergek et al., 2008). The emphasis is on understanding innovation and diffusion processes and on making diagnoses indicating which system problem the policy efforts should ideally focus on. This lens can also be used to see whether positive developments have taken place at a certain time interval in the system aspects in which improvement is most needed. From the effect ladder perspective, this analysis is about demonstrating correlations or showing that the most necessary improvements have not taken place.

Although evaluation is discussed infrequently in this literature, there are examples of studies that have this structure. See, for example, the evaluation of the Top Sector approach discussed in Box 3.2 and the recent evaluation of the comparable Australian Industrial Growth Centre Initiative. The Swedish evaluation discussed in perspective B also used the TIS framework in the evaluations, as is also shown in Figure 4.4 (Åström et al., 2022). The same applies to the evaluations of the NWO Take-off programme and the Dutch variant of the SBIR; here too, the system effects of the implemented policy were examined on the basis of the TIS key processes (Dialogic, 2017).

Figure 4.4 Example of analysis of the extent to which TIS key processes have been strengthened by policy. Source: Arnold (2022)

Function	Current practice	Potential with SIP 1.0
TIS functions	(Scale: 1 = low; 5 = high)	
Entrepreneurial experimentation with new technologies, markets, and business opportunities	2	3
Knowledge development, via R&D and learning-by-doing	3	4
Knowledge diffusion through networks	3	4
Directionality	2	3
Market formation by opening market space or articulating demand	2	2
Market formation by creating protected space for niche innovations	1	1
Legitimation	3	4
Resource mobilisation	1	1
Developing positive externalities	2	2
Other change agency functions		
Creating arenas for priority setting	2	3
Building actor networks or coalitions	3	4
Developing guiding visions	1	4
Action at the political and policy levels	1	4
'Creative destruction', phase-out management	1	1
Reflexivity	2	4

E. Structural change

Policy changes how a system works. The question is what an S/T programme seeks to bring about. This may be, for example, the creation of policy support to strengthen innovation systems benefiting specific innovation paths. For effect measurement with a view to structural change, an important point is whether those systems contribute to the success of a selection of innovation paths. Figure 4.5 shows how an evaluation structure could look. If the policy goal is to bring more focus and acceleration to specific innovation paths, effect measurement should focus on the following questions:

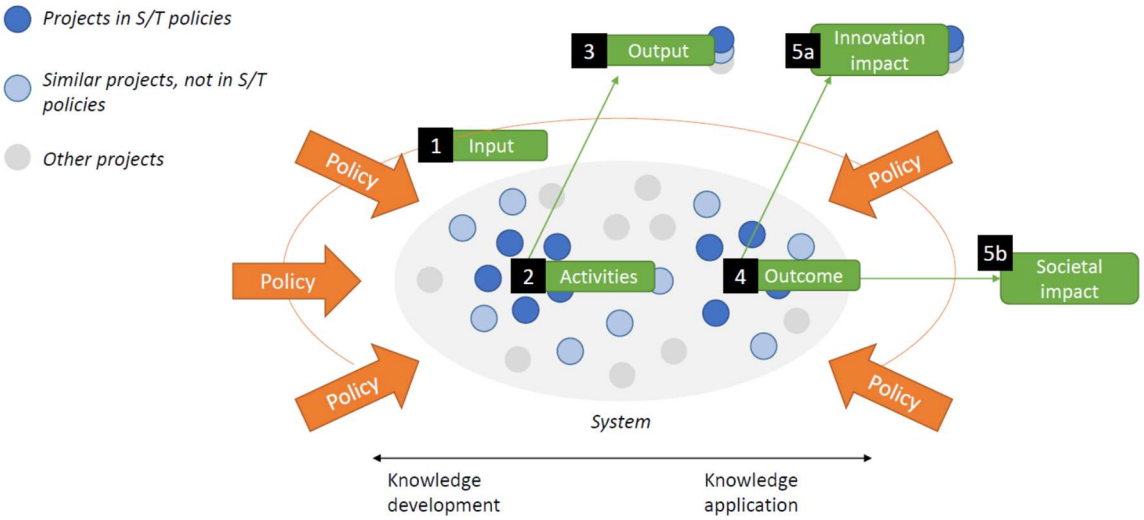
1. Input: Is there more thematic and organisational and administrative coherence between potentially complementary policy instruments and efforts?
2. Activities: Is there thematic convergence in the innovation projects supported by various schemes (in relation to the prioritised themes)?
3. Output: Are the projects relating to prioritised themes developing faster and more successfully (e.g. in terms of patents, publications, prototypes) than other innovation projects?
4. Outcome: Do we see more innovation activities also arising *outside* the schemes focused on the application of solutions as found *in* the schemes?

5a. Innovation impact: Are the innovation activities referred to under 4. also associated with better economic performances by organisations that undertake these activities?

5b. Societal impact: Are the innovations supported by the S/T policy also applied and scaled up relatively faster?

The key principle in these questions is always that remedying a failure, such as a lack of direction, should primarily involve creating more momentum for specific innovation directions (operationalised as greater cohesion in innovation efforts), because of the connecting coordination activities and the targeted policy incentives. One question is whether more attention will be paid to innovation directions within the schemes and whether the policy incentives and innovation projects will initiate so much (i.e. achieve system strengthening) that organisations will join and invest in them.

Figure 4.5. Effect measurement structure for various types of objective (1-5), associated with the S/T policy to target innovation



One way of observing this process is by looking at the extent to which company-specific innovation projects, for example in the WBSO scheme, conform to these directions. Since the WBSO is a relatively generic innovation scheme, without detailed control, the effect of S/T policy could be apparent from an observed increased willingness on the part of businesses to perform R&D in the field of products and processes that fit in with the policy-supported innovation paths. If such an increase is not discernible, it is unlikely that the policy will be effective. If the increase is discernible, it is necessary to investigate what the contribution from policy will be. By extension, it is necessary to investigate to what extent complementary economic performances are important (see effect 5a Innovation impact).

A requirement for this evaluation approach is that it must be clear what the solution directions are and how innovation projects within the associated schemes and especially outside them relate to those directions. Projects must be labelled accurately for the comparison, because labelling makes it possible to see whether the chosen innovation directions are developing faster and better than the non-chosen directions. A promising point in this regard is the principle that S/T policy should set clear directions for the combining of efforts. At the time of the evaluation, it should be determined to what extent these directions are clear and what directions are involved.

A limitation of this analysis design is that it ignores the fact that it is not the case in all S/T programmes that the relevant solutions result from system change or transitions. In addition, the analyses do not explicitly look at changes in demand for new products/processes/services. A larger number of innovation projects is still no guarantee that those innovations will be embraced and will be effective in achieving societal objectives.

F. Societal impact

The final evaluation perspective involves examining developments in the target variables that S/T policy is ultimately all about.

Changes in target indicators are often used to make statements about efficiency. People look, for example, at the reduction in CO_2 (or CO_2 equivalent) per euro of policy cost. In order to make a genuine statement about the effectiveness and efficiency of policy, however, it is necessary to know to what extent performances are due to the impulses provided. They may also be due to exogenous

developments.[15] Conversely, it is also possible that the policy fits together well in terms of logic, governance, policy mix and even system changes, but that unforeseen exogenous developments mean that this is not translated into the desired performances. As described earlier, establishing causality in S/T policies is complicated because of the multitude of synergistic policy incentives that also have an effect individually and because of other developments that together influence the intended objective. A second problem is that the effect may not be linear in the policy costs and over time. When assessing performances on target variables, it is therefore important to devote attention to the expected course, with uncertainty also being taken into account in the policy. For example, performances may initially develop positively but then stagnate. Policy then bears fruit in the short term, but is probably less effective and efficient in the longer term.[16] The reverse can also be the case; no direct results will be visible in the short term, but they will be in the long term.

In summary, when evaluating societal impact, it is important to establish reliably to what extent developments in monitored target indicators are the result of implemented policy. If that is not possible, observed performances can sometimes be compared to situations elsewhere. In such comparisons – if an underlying causal analysis is not possible – evaluators must be cautious in attributing observed differences to the conducted policy.

4.5. Selection and application of the evaluation framework perspectives

Considerations with regard to the choice of perspective

Every perspective in the proposed evaluation framework highlights a different aspect of the operation and effectiveness of S/T policy. The 'higher' perspectives (D-F) lend themselves best to summative analyses with a view to accountability for the changes made, while the emphasis in the 'lower' perspectives A-C is on formative research into the substantiation and implementation of the policy. Strictly speaking, the latter category of perspectives generally does not belong to the field of effect measurements. As described in Chapter 3, S/T policies sometimes have limited potential to isolate the effect of programmes which, through better coordination and policy coordination, seek to boost systems that are ultimately important for achieving a socially desirable objective. In those cases, it becomes more relevant to be able to make statements about successes and failures in the organisation of policy that is at least able to address essential system/transition issues (even if we cannot offer a clear causal determination of how that will subsequently turn out). This is particularly true if S/T policies themselves provide few clearly identifiable incentives for 'the field', and instead mainly serve to strengthen the connection between existing (complementary) policy initiatives. The extent to which the policy is effective can then be examined by looking for improvements in the coherence of policy initiatives. If such coherence exists, its effect can also be determined with perspectives D-F. If the coherence does not exist at the outset, an examination of the effect is immediately less relevant.

The above logic stands or falls on the extent to which there are principles that 'coherent policy' must fulfil. In theory, it is possible that the improvement of the connection between policy initiatives will fail, but that progress will nevertheless be made with a transition for unexpected reasons. This possibility underlines not only the maxim that one should at any rate try to apply perspectives D-F,

[15] It is conceivable that policy aimed at societal challenges will sometimes link up with developments that already have a certain momentum, and which therefore also generate the social pressure that brings the policy into existence. In that case, after a few years there may be a correlation between policy Introduction and performance in target indicators, without the latter necessarily being a consequence of policy. It is also possible that the policy will have a negative influence, for example when newly imposed coordination structures impede existing coordination mechanisms. This scenario underlines the importance of an incisive analysis of the intervention rationale, as it should reveal whether there really are problems that require policy intervention (see perspective A).

[16] For an example, see the PBL report “Major tasks in limited space” (April 2021), which warns that hundreds of millions invested to bring about changes in agriculture could be thrown overboard in the medium term.

but also the importance of follow-up research into well-functioning governance processes and policy mixes: the more we know about how S/T policies should be coordinated, the better we can assess concrete cases on that basis.

In general terms, with S/T policies it is desirable to evaluate as many perspectives as possible, since it is precisely the consistency between subanalyses that indicates the level of effectiveness of policy: for example, are there already problems in strengthening a policy mix, is the system changing properly, and are there problems with bringing about structural changes (possibly due to exogenous factors)?[17] Findings from various sub-analyses can give cause to conduct targeted in-depth analysis: if a particular aspect of system change excels or lags behind, can this be attributed to the characteristics of the policy design and implementation?

Of course, for practical reasons it is not always possible to cover all perspectives in a single evaluation study. If a choice has to be made, it is sensible to take at least two points into account. The first is the nature of the policy, whereby programmes with an emphasis on coordination and attuning of the policy mix (rather than providing new incentives) benefit relatively more from insights based on perspectives A-C. The policy objective of a programme such as the Mission-oriented Top Sector and Innovation Policy is to create more synergy in various resources and initiatives that can be used to contribute to societal challenges. The first question to answer is whether there is indeed more momentum, which can be evidenced among other things by consistency in the criteria of various measures, a focus on the subjects to which actors direct their own investments, or networking around those subjects (see next section). Investigating the ability to adapt policy to problems is more important as those problems shift, which may often be the case, especially in transitions. When S/T policy itself also provides concrete incentives, as in the case of climate policy, it becomes more relevant to also look closely at the influence those incentives have on the basis of perspectives D-F. The second factor to take into account is the phase of the policy cycle. If the Periodic Evaluation Scheme requires an accountability investigation to be carried out every five years, that automatically points to perspectives D-F, while for perspectives A-C the natural choice is to monitor on a continuous basis which policy adjustments are being made and what reactions will follow. Since coordination-intensive S/T policy often relies on working the 'margins' of a large number of policy activities and initiatives (e.g. decisions on subjects that are given extra or less weight), it may be sensible to reflect continuously on who and what is being mobilised, not only because action can then be taken earlier, but also because it is not easy to build a retrospective picture of the accumulation of all those small adjustments.

Operationalising the perspectives

In order to apply the perspectives from the evaluation framework in practice, it is necessary to operationalise them. In the case of traditional innovation policy, this is often a matter of looking for indicators for objectives such as 'increasing R&D intensity/collaboration'. Once the indicators are known, the next step is identifying a data source with variables that can be used to measure the indicator.

R&D intensity in the Netherlands is often approached by looking at a company's WBSO hours (on which Statistics Netherlands holds confidential information), by using a company's answers to relevant questions in the R&D survey / Community Innovation Survey (also known at Statistics Netherlands, for a selection of companies), or by having evaluators conduct a targeted survey.

In the case of S/T policies, the step towards indicators (let alone data sources) is more complex. In the first place, there is a greater variety of objectives for which indicators are appropriate, because the

[17] This logic of tracing the point up to which there are desirable changes in policy theory is consistent with the principles of the process tracing methodology; see also Chapter 5.

ambitions can relate to systems and transitions in a large number of areas. In addition, such objectives are much more abstract than increasing R&D investments or their direct results. The term 'system' can refer to innovation systems, production-consumption systems (what is produced and purchased, and by whom) or socio-technical systems. In the case of the latter interpretation particularly, it is difficult to get a good grasp of it with data. The ultimate transition to which S/T policies is intended to contribute can sometimes be monitored effectively (e.g. the amount of CO_2 emission reduction in a particular sector), but if this cannot be attributed to the policy, it will be necessary to resort to indicators for intermediate goals. It is then essential to properly understand the mechanisms through which the ultimate objective should be achieved, and thus what the relevant intermediate goals are in order to set up the evaluation. A third complication, which has already been discussed in detail, is that in the case of S/T policies it is often impossible to deduce that effect from the official policy documents. In that case it is first necessary to learn (about the policy) before it is sensible to start measuring.

Consequently, operationalising the evaluation framework perspectives starts by asking the right questions. The answers to these questions are not quantitatively measurable indicators, but they are indications of the state of the policy. Table 4.2 provides some sample questions for each perspective. It this also provides a ready summary of what the various perspectives have to offer.

Table 4.2. Overview of sample questions for each perspective from the proposed evaluation framework

Perspective	**Examples of questions**
A: Intervention rationale	What are the types of failure that require S/T policies (i.e. with a focus on coordination and streamlining)? Where precisely are the issues, and what is the evidence for them? For example, without the policy, are there insufficient possibilities and policy options to contribute to system change and/or transitions, and why are the market and cooperating actors unable to find a solution themselves?
B: Governance process and policy mix	Are there principles indicating what the coordination in the S/T policies should focus on, and does the coordination comply with those principles? Is there commitment from the relevant parties? Does the coordination lead to appropriate adjustments in the policy mix (i.e. modifications that align system changes and resulting activities better with policy objectives)? To what extent is this reflected in the development or phasing out of schemes and in adjustments to the financing and design of schemes?
C: Policy vs issues	What are the issues (in terms of system processes or structures) that complicate system change or make transition processes more difficult, and how are S/T programmes and the policy instruments to which that policy applies responding to these? Is most energy devoted to the biggest issues, and what results are achieved by strengthening the weakest aspects of the system?
D: System change	Do we see improvements in the system processes that enable the development and dissemination of innovations? Is there more development/exchange of knowledge, market formation, etc. (depending on which process was weak) around desirable innovations?
E: Structural change	Do innovation systems / socio-technical / production-consumption systems change structurally? This concerns the transformation of entire systems, including the parts that fall outside the narrow scope of the schemes associated with the S/T policies. An initial indication would be whether they show more innovation activity.
F: Societal impact	Are the interim goals and ultimate objectives achieved? And to what extent is that the result of system and other changes attributable to the influence of the S/T policies? For example, if there are sectors in which CO_2 savings have been achieved, is that also due to innovations that have been demonstrably boosted by the policy?

Since the sample questions have to cover different types of S/T policies they are formulated in general terms. In order to make the questions as a whole somewhat more specific, we provide a concise illustration based on one of the cases that also arise in Chapter 8: the Mission-oriented Top Sector and Innovation Policy (MTIB).

- Perspective A: Assessing the intervention rationale for the MTIB requires an answer to the question of why setting up coordination structures such as theme teams and mission teams, including their

mandates for the creation of the Knowledge and Innovation Covenant (KIC), would be an appropriate way to give an overall positive boost to innovation for the societal objectives embodied in the 25 missions. From the scientific literature, it is known that markets sometimes fail to make synergistic investments in new solution/transition paths, and that different governments can obstruct each other when providing support for these. It is therefore easy to argue at the outset that 'innovation for transitions' requires government intervention on the basis of transformation failures such as directionality failure, reflexivity failure and policy coordination failure. The question, however, is whether such forms of failure also actually apply to all 25 missions, or to the four mission themes under which they fall. It is possible that there are also all kinds of forces that automatically mean that, on the basis of their considerations, market participants and other stakeholders (including governments) demonstrate behaviour that contributes to a mission objective such as “Making the security profession one of the 10 most attractive professions in the Netherlands by 2030”. In this sense, it is appropriate to verify for each of the mission objectives (e.g. by means of desk study and interviews with policymakers, domain actors and domain experts) what the precise grounds are for assuming that an S/T policy such as the MTIB is an appropriate policy strategy. In addition to reviewing possible transformation failures, it is also important to look at underlying market and system failures: if it appears true that convergence around a socially desirable solution does not arise automatically, the follow-up question is why that is the case. The answers to this could differ substantially for the 25 missions, illustrating how difficult it can be to speak of overall quality (or in the case of other perspectives: the effect) of S/T policies.

- Perspective B: In view of the character of the MTIB – creating momentum by streamlining a policy mix and mobilising/uniting a range of actors – it is natural to devote extensive attention to the establishment of governance processes and structures. Since this requires the necessary knowledge of the MTIB, the list of relevant sample questions has been included in the case study in Section 8.1. These questions are specific versions of the examples in Table 4.2, with the common denominator being attention devoted to thematic coherence in policy, agendas and innovation/diffusion efforts that the MTIB is intended to bring about. In addition, it is also possible to analyse the extent to which the MTIB is actually taken on board by the parties who are represented in the governance structures, and who, on behalf of their rank and file, help to expose and tackle issues. This can be evidenced, for example, by commitment in the form of time and resources that parties invest in participating in governance activities. Whether the captains of industry / science (etc.) continue to participate in the theme team meetings indicates the likelihood that the MTIB will contribute to mobilising and coordinating the actions of relevant stakeholders.

- Perspectives C&D: To determine whether the coordination activities in the MTIB also set the right priorities, it is important to link these to the system issues that arise in the themes that are central in the MTIB programming. Knowledge of those issues is also essential for studying system changes. Building on the evaluation of the Top Sectors approach, in both cases it is possible to opt for the use of the dimensions and indicators associated with the Technological Innovation Systems (TIS) framework. As discussed in Section 4.4, this framework distinguishes a number of key processes or 'functions' that are important for the proper functioning of an innovation system focused on a specific technology. The difficulty with societal mission objectives, however, is that it is not possible to designate a single technology in advance, and that achieving those objectives also requires a different type of intervention. For better alignment with such a context, which is characteristic of S/T policies in a broader sense, work is currently being carried out on the conceptualisation and operationalisation of 'mission-oriented innovation systems’ (MIS) – see Elzinga et al. (2020). Table 4.3 on the next page shows the MIS functions and some examples of underlying indicators. In practice, the main challenge is to define clearly the indicators that should be measured for each system/theme.

Table 4.3. Examples of indicators for the analysis of mission-oriented innovation systems (Elzinga et al., 2020).

Function	Examples of indicators
Experimentation by companies	Number of new and existing business operators experimenting with relevant solutions Number of projects started/stopped Number of businesses started/stopped
Knowledge development	Number of (scientific and professional) publications Number of knowledge projects started/stopped Impact of publications (citations, use in society) Number of new research groups Number of new research programmes Number of new chairs Number of new patents
Knowledge dissemination	Number of symposia and conferences Number of knowledge networks Number of joint publications Number of knowledge-sharing activities between different actors (businesses, consumers, knowledge institutions, governments)
Sharpening the problem focus	Visibility of the challenge in media Visibility of the challenge in business strategy Visibility of the challenge in industry organisation strategy Visibility of the strategy in research programmes Extent to which challenge is on the agenda of relevant organisations (priority) Specificity of the descriptions of the challenge in strategic documents Stimulating laws and regulations with regard to generic challenge
Sharpening the solution focus	Statement of expectations for specific solutions Social discourse in media around specific solutions Legislation and regulations encouraging a specific solution
Market creation	Business activities that persuade the consumer to adopt Regulations requiring solutions to be applied Fiscal or subsidy instruments that make solutions financially more attractive Standards / labels that make solutions distinctive and encourage / require adoption
Mobilising resources	Provision of financial resources for innovation, research, pilots, investments by governments, businesses and other organisations Training and provision of personnel with the right knowledge and capability Provision of the required infrastructure and raw materials
Creating legitimacy	Lobbying and media statements intended to reduce fear / resistance around new projects Balance in lobbying activities for and against the mission and associated solutions
Coordination	Formation of coalitions around the transition Emergence of leading organisations to bring the field together Strategic documents that compare / analyse different solution directions Emergence of coherent visions of the transition
Putting pressure on the current regime	Regulations stating that existing activities need to be scaled back. Abolition of subsidies / advantages for existing activities Voluntary agreements with market participants to change course Partnership discourse / activism against current practices Actions aimed at naming 'wrong' consumer behaviour

- Perspective E: The perspective concerning structural changes has no general conceptual basis. In the case of the MTIB, it is natural to consider at least changes in cooperative structures, or R&D and innovation networks. For the original Top Sectors approach, after the formal evaluation a study was conducted to determine whether those networks changed after the Top Sectors and TKIs were introduced – see Janssen et al. (2020). This showed among other things that parties started collaborating over greater geographic distances, and that parties belonging to the ecosystem of the same TKI started collaborating over greater cognitive distances (combining knowledge that was normally combined less often). In the same way it is now also possible to see whether the introduction of the MTIB has brought about more changes. An initial analysis based on public-private R&D projects shows that parties that collaborate on a project relating to the mission themes of Energy Transition & Sustainability or Health & Care generally bridge shorter cognitive distances compared to other public-private R&D projects – see Janssen and Abbasiharofteh (2022).

The question is how this will develop over the years, and whether there will be increasing and more diverse cooperation on the solution directions that are central to the MTIB programming (compared to similar but non-selected directions).

- Perspective F: Ultimately, the MTIB aims to contribute to achieving the mission objectives. The most direct policy objective, however, is that the MTIB will accelerate the development and application of innovations that could contribute to those goals. This can be measured by the extent to which private behaviour far beyond the MTIB also conforms to the priorities in the MTIB programming. When it comes to innovation, it is possible to look automatically at the descriptions of individual companies' WBSO projects, but there may also be other suitable data sources for each mission. See again the case study in Chapter 8 for more detail on this point.

5. Methods for evaluating system and transition policies

Once the evaluation framework has helped determine which policy aspects should be examined, the next question is which research methods are most appropriate. Section 5.1 provides an overview of various options and Section 5.2 continues with a discussion of their feasibility for evaluating S/T policies.

5.1. Overview of methods

This section provides an overview of empirical research methods. It uses the inventory and description of research methods in the online Policy Evaluations Toolbox. Monitoring is not included as a method, since the recording of developments in relevant indicators serves to feed other (possibly more evaluative) methods. Policy monitoring is a necessary precondition and a means of collecting relevant data for evaluation.

Effectiveness

Table 5.1 shows which methods appear most suitable for examining the effectiveness of S/T policies.

- The methods can be classified on the basis of *the type of empirical data* required for application. In two cases the methods are mainly theoretical[18] in nature: systematic reviews and meta-analyses and simulations. Meta-analyses and simulations are quantitative, whereas systematic reviews are qualitative. Subsequently, there are various qualitative and quantitative methods, each of which has a different application.
- In order to make this clear, the table also shows the *analysis objectives* to which the methods lend themselves. At least three analysis objectives are important for effectiveness evaluations: explaining mechanisms (the ways in which policy and possibly other factors/conditions contribute to the achievement of objectives), determining causality (to what extent it can be determined whether and when A leads to B) and measuring the (usually average) size of the effect of a policy intervention.

Table 5.1. Characteristics of research methods with relevance for the evaluation of effectiveness

	Data			**Analysis objective**		
Method	Theoretical	Qualitative	Quantitative	Identifying operating mechanisms	Determining causality	Measuring size of effect
Systematic reviews & meta-analysis	X		X	X		
Simulations	X			X		
Reflexive evaluation		X		X		
Outcome harvesting		X		X		
Case studies		X		X	X	
Qualitative Comparative Analysis (QCA)		X		X		
Contribution analysis		X	X	X		
Process tracing		X	X		X	
Regression analysis			X			X
Experimental/quasi-experimental methods			X		X	X

[18] Theoretical means that these analyses are based not on data generated in the system to be evaluated, but on models or on analyses conducted in a different context.

Compared to simulations, systematic reviews and meta-analyses have a different analysis objective. Consulting existing literature can help to better understand the normative principles governing policy. In addition, analysing existing studies is a way of demonstrating the extent to which policy effects arise, provided there are of course previous studies of the same kind of policy and the same policy objectives. This does not assess the causality of S/T policies, but as more relevant previous research becomes available it is easier to estimate and reason how likely it is that the current S/T policy is having a causal effect. Simulations, on the other hand, are better for exploring mechanisms, especially when it comes to the interplay of behaviours that are known individually but not collectively.

There are also a number of qualitative methods that can be used to determine effectiveness. The nature and applicability of these qualitative methods differ in a number of respects.

- *Reflexive evaluation* is a means of investigating, together with policymakers and stakeholders, how policy relates to (often complex) societal issues. In the first place, this involves gaining a better understanding of the operation attributed to selected governance structures, policy actors and policy interventions, in order to construct (or reconstruct) how the policy works and which issues remain or emerge. This method helps to explain the mechanisms through which policy can contribute to the achievement of the intended objective, but does not say whether that will actually happen (or to what extent);
- Outcome harvesting is (also) a method for assessing operating mechanisms. In contrast to reflexive evaluation, it is centred not on the policy itself but on the outcome of the policy. By working backwards from the outcome of that policy to the (causal) chain that was followed, or in other words by looking at the factors that explain how an observed outcome occurs, it is possible to find out what the relative role of the policy is relative to other factors of importance for the outcome;
- Case studies are a combination of a research method and a data collection method and can be used for many purposes. By examining the role of the policy conducted in one or more cases (projects, missions or transitions), it is possible to gain a clearer picture of the mechanisms that generate outcomes. For the cases studied, it is sometimes possible to determine whether there is a causal relationship between policy and outcome, but in case studies it is often impossible to measure this effect quantitatively and generalise it to other cases (lack of external validity);
- Qualitative Comparative Analysis (QCA) is a method for systematically comparing a number of cases to determine which factors (or combination of factors) is associated with a particular outcome. A precondition for this method is that the cases are varied in terms of the performance for that outcome and that in all cases it is known whether certain conditions are in place. These conditions may include the presence or absence of policy (or a particular aspect of policy), but it is also possible that policy is in place in all the cases studied. A condition that does not vary across cases will not be included in the analysis (because it has no explanatory value indicating differences between the cases). When there is variation, the QCA method helps to find a correlation between a configuration of conditions and whether or not a policy objective is achieved, but it is not possible to say anything about the role of the policy itself;
- Contribution analysis & process tracing are methods for illustrating causal relationships. As with reflexive evaluation, existing policy is the starting point for the effect measurement. Both methods essentially comprise the collection of qualitative and/or quantitative data with which the existing policy theory conducted to determine whether assumed changes and impacts are present. This does not show the quantitative size of a causal effect, but it does reveal the extent to which there are developments that are consistent with the goal of the policy theory. In contrast to contribution analysis, process tracing involves the testing of hypotheses based on the policy theory.

The main difference between the two quantitative methods in Table 5.1 is that regression analyses are to be used primarily to measure correlations between the policy impulse and the target range (and to correct as fully as possible for other factors that affect the target variables), while experimental or quasi-experimental methods are characterised by their potential to determine causality as reliably as possible.

The challenge in the case of efficiency evaluations is generally to draw conclusions that are relevant to the ultimate policy objectives pursued, while also providing a valid picture of causal connections. When evaluating S/T policies it is often difficult to achieve both ambitions with a single method and combining multiple methods is an effective solution. In such mixed-methods analyses, the various methods can relate to each other in different ways, and that also determines whether they should be deployed simultaneously or sequentially. Methods can thus be combined in order to triangulate observations and thereby maintain the reliability of findings. Formative methods can also be used to gain an initial view of operating mechanisms and on that basis to design a summative effect measurement.

Efficiency

Evaluations that aim to draw conclusions about efficiency will benefit from a number of supplementary research methods (Table 5.2). Methods that can be used to measure technical efficiency usually belong to the family of productivity research methods – such as data envelope analysis (DEA), stochastic frontier analysis (SFA) and benchmarking (e.g. comparisons with investments in similar S/T policies in other countries and the developments seen there in terms of output and outcome variables). These methods are aimed at determining and comparing observed performances and/or costs, and possibly the relationship between them.

Table 5.2. Characteristics of research methods relevant to the evaluation of efficiency

	Data			**Analysis objective**		
Method	Theoretical	Qualitative	Quantitative	Measuring technical efficiency	Measuring allocative efficiency	Measuring dynamic efficiency
Simulations	X		X	X		
DEA			X	X		
Stochastic frontier analysis			X	X		
Benchmarking			X	X		X
SCBA			X			X
Case studies		X		X	X	X

Qualitative methods are also ideal for studying dynamic efficiency, because they require a number of processes to be visible. Examples are the extent to which decisions are divided into steps, the extent to which experiments are conducted and the extent to which an option to defer policy is included. Social cost-benefit analyses (SCBAs) can also be used to weigh the intended future effects against current efforts, if empirical estimates of expected effects can be made based on known elasticities or models, for example. This provides an *ex ante* picture of the overall policy effort that may change in certain respects over time. It also indicates which mechanisms envisaged by the policy have unknown effects and where effect measurement is therefore necessary. In addition, it is possible to use SCBAs not only *ex ante* but also in an *ex post* evaluation. This shows whether the advantages (benefits) of the policy outweigh the disadvantages (costs). With an "SCBA as a conceptual framework", it is possible to determine which effects are unknown and where effect measurement is therefore necessary.

Allocative efficiency is difficult to gauge in practice, because it requires an assessment of the use of resources relative to other policy fields. Case studies appear most suitable for this, because It is difficult to make a comprehensive comparison of policy fields, although lessons can be drawn from cases about the efficiency of the contribution from policy efforts.

5.2. Descriptions and application of individual methods

In this section we briefly discuss some of the research methods. For each method there is a discussion of how it can be applied in the context of S/T policies. More detailed descriptions and more general commentaries on the nature, application, requirements, limits, costs and lead times of the methods can be found on the website of the Policy Evaluations Toolbox.

Systematic reviews & meta-analysis

Description

A systematic review is a specific form of literature research in which the results of a number of scientific articles on a specific subject are summarised and assessed on a targeted basis. A meta-analysis is an extension of the systematic review. It is a static method used to compare and combine the quantitative results of multiple studies.

Relevance to S/T policies

A systematic review can be useful in examining whether there is existing empirical evidence and/or a theoretical basis for the existence of certain issues, and to identify operating mechanisms through which specific policy approaches can be adopted to remedy those issues. Examining this amounts to checking the credibility of the policy theory. This method is suitable for mirroring the intervention logic (perspective A) and the match between policy and issues (perspective C) to scientific insights. Systematic reviews can also be used to develop frameworks (of a normative nature) that state the principles that should underpin the transition governance (perspective B). This is particularly helpful when establishing a policy theory. Meta-analyses can be used to determine the average size of effects found in the literature. This gives an impression of the possible effects of the policy to be evaluated.

Simulations

Description

In a simulation, researchers imitate the reality by simplifying it in a model. That model describes a number of rules which the individual actors (agents) in the model (people or businesses) must comply with. These actors then 'play' the 'game' according to the rules and produce a specific outcome, such as a shared consensus on a particular problem or an equilibrium in a market in the case of economic models. By changing the rules of play it is possible to study what influence certain factors can have on the outcomes. Simulation is therefore an ideal method for learning, both about the behaviour of actors and about the likely impact of interventions on that behaviour. There are two types of simulations that use people (game simulations) or computational models (computer simulations) respectively. The limitations of such simulations lie in model assumptions and in the extent to which a laboratory environment represents reality.

Relevance to S/T policies

S/T policies by definition address systems. A characteristic of many systems is their complexity resulting from a non-linear feedback loop and a concurrence of impulses or behaviours. Examples

include technological or other innovation systems that can only properly promote the development and diffusion of a socially desirable innovation if a series of complementary key processes has been put in place. As long as processes remain underdeveloped, a development path can only get under way slowly, whereas rectifying the most underdeveloped process can suddenly trigger a major acceleration. Something similar can also arise in the development of complementary technologies (in energy and elsewhere) that are interdependent and hence can impede but also accelerate each other's development. In this kind of situation it can be valuable to investigate *ex ante* how the concurrence of familiar individual dynamics and behaviours would work. For example, in computer simulations it is possible to analyse the relative influence of a single transition factor in a whole set of transition factors, in order to make a statement about the probability that structural changes and societal impact will arise in a particular policy incentive (perspectives E and F). An alternative is to initiate a combination of behaviours with game simulations. This is also a means of developing future scenarios that can support policymaking.

Reflexive evaluation

Description

Reflexive evaluation describes an approach to evaluation inspired by methods discussed in Section 4.1, such as developmental innovation, whose key characteristics include participation by policymakers themselves. In this type of evaluation, evaluators and policy stakeholders work together to use the evaluation research to provide useful and meaningful insights for policy development and/or implementation (Folkert et al., 2018). Since policy research and policy development are interlinked, this method can best be used in the *ex durante* phase, for example to assess policy impulses in terms of their contribution to long-term objectives such as a transition. Reflexive evaluation is primarily a method for generating knowledge about which policy instruments will work and when (where and under what circumstances), but it provides no causal evidence of effectiveness. Hence it is often also a form of 'realistic' evaluation: despite the lack of causal evidence, it is still useful for a better understanding of what this (unmeasured) effectiveness depends on. Although reflexive evaluation is mainly used formatively, summative elements can also be introduced; accountability is possible as long as evaluators maintain sufficient independence in the design and interpretation of the research.[19]

An example of reflexive evaluation is the role of researchers at the Fraunhofer ISI Institute for German mission policy. These researchers are continuously engaged with their own analyses and interactions with all kinds of stakeholders in the German innovation system, and occasionally publish analyses of these (we refer to some of those interim evaluations elsewhere in the report). They have an explicit role as a 'thorn in the side'; by means of their questions and analyses they force policymakers to stay focused on arguing and demonstrating the relationship between policy (design and implementation) and outcome. The researchers' continuous critical thinking differs from the model in which governments commission studies at specific times by external researchers who are relatively remote from policy.

An advantage of reflexive evaluation is that the implementing researchers follow the development and implementation of policy over a longer period of time, and already have a good impression of its complexity before they focus on more specific analyses. This is in contrast to external researchers who are asked to investigate effectiveness within a few months but in practice first have to devote a lot of time and capacity to understanding the essential details. A risk of this model is that the evaluation and

[19] Accountability takes on a broader meaning in a reflexive evaluation setting. In addition to the 'regular' meaning of accountability for effectiveness and efficiency, reflexive evaluation is particularly suitable for horizontal accountability (between the parties involved) and internal accountability (policy consistency) for policy (Regeer et al. 2016).

policy process are seen as one and the same thing, whereas separation is important to guarantee the quality and credibility of evaluation.

Relevance to S/T policies

Reflexive evaluation is useful when the operation of the S/T policy is so complex that the underlying policy theory cannot be properly deduced from existing policy documents. An understanding of how the policy should ultimately have an impact can then benefit from combining the evaluation knowledge of researchers (e.g. knowledge of developing a policy theory and identifying related indicators) with the expertise of policy stakeholders. The method is therefore suitable if it is unclear through which mechanisms any effects will be achieved, and there is a need to structure relationships between inputs and effects of various kinds (perspective A, intervention logic). Reflexive evaluation thus also provides a basis that can be used in the other perspectives to assess policy and outcomes.

The dynamics of a transition also require a learning attitude, so that further adjustments can be made to focus policy on the issues prevailing at that time. Reflexive evaluation can be part of managing a transition; there is a continuous assessment and accountability element that strengthens the policy structure and implementation (perspectives B and C). During the evaluation process these insights can lead to an adjustment.

Due to the focus on how policy is implemented, and what direct results this delivers in the light of a long-term objective or ambition, reflexive evaluation enables conclusions to be drawn about the effect of S/T policy. The method therefore focuses on transition steps that are close in time, with subsequent steps possibly requiring policy change – and hence reflexivity. This is mainly about highlighting arguments that explain how the pursued policy affects 'second-order' effect categories such as system strengthening and system transformation (perspective D). The method is appropriate for making statements about the feasibility of transition objectives that lie in the distant future (20-50 years).

Case studies

Description

A case study is a research strategy in which one or more cases are studied in depth in their natural environment, often in a delimited spatial and temporal context. In policy research, case studies are often used for complex decision-making processes, for example. A case study is characterised by a small number of research units, a labour-intensive approach, a preference for depth over breadth of research and a reasoned, informed choice of cases. A case study can be descriptive, testing or explanatory, but the external validity of the findings is often limited.

Relevance to S/T policies

In the context of S/T policies, it is not obvious what the term ‘case’ precisely relates to: it may concern, *inter alia,* organisations, projects, agendas, instruments, innovation systems, partial missions or transition paths, for example. A potentially useful definition is that of an innovation or solution path, which, for example, can revolve around a technological innovation and related business model innovations. To understand the importance of policy for an innovation such as ‘electric flight’, it is sometimes helpful to step back from specific actors and projects and reconstruct at the innovation level the developments that were important for growth or stagnation. A technique such as history event mapping is suitable for gaining a structured insight into events over time in relation to the innovation, and which causal links can be made (for the case in question) with the policy.

This method can provide a picture of the issues that the innovation has had to deal with (perspective C), or which system conditions were beneficial or a hindrance (and whether policy influenced this;

perspective D). Such an analysis can also be performed at project level, but this is often less meaningful, because projects are strongly associated with individual schemes (which in turn have their own scope, e.g. in terms of the development phase). For S/T policies, it is the meso level of systems that is important; they must be established in such a way that innovations can pass through all development phases through to large-scale adoption. Case studies should therefore perhaps focus on concrete innovations, which are of course often based on all kinds of underlying projects.

Outcome harvesting[20]

Description

Outcome harvesting (OH) is a method for studying in complex situations how a certain outcome results from a chain of causes (interim outcomes). A distinctive characteristic of OH is the open nature of the method. It is not based on a predefined policy theory, but defines empirically identified changes as an 'outcome' (e.g. growth in circular practices) and then tries to ascertain how an intervention (e.g. a sudden change in a policy measure) has contributed to this specific change. The Most Significant Change (MSC) method is a concise variant of OH in which evidence is collected at case level on factors contributing to an outcome. The focus of both OH and MSC is on the inventory of outcomes, which can be useful if inputs, activities, outputs and results are not specifically identified and measurable in advance. An important difference as compared to monitoring is that the latter method requires knowledge of which indicators need to be measured in order to keep track of the policy effort and target range.

Relevance to S/T policies

Accurate policy theory is often unavailable for complex policy programmes. Specifically for S/T policies, this is also partly inherent in the fact that, in contrast to conventional innovation policy instruments, it is often impossible initially to identify a concrete problem that explains why a socially desirable objective is not being achieved. It is the actual pursuit of activity that should then reveal what is holding back innovation or even transition, and how policy can respond. In that sense OH is potentially useful in the evaluation of S/T policies. Once programmes and instruments have been set up to boost system strengthening or transformation, research can be conducted into the ways in which the impulse provided relates to the actual performances. For example, facilitating coordination is often a central element of S/T policies, but evaluations must show how that coordination ultimately leads to impact. This impact may lie, for example, in new R&D partnerships (private-public, private-private), coordination between businesses and governments, streamlining of development and adoption etc. The OH method, which provides a general insight into intended and unintended outcomes, therefore fits in well with the principle that with S/T policy it is not always possible to specify in advance how the policy will lead to desirable changes.

Since OH is applied by examining how successful projects have benefited, the method does not reveal more structural changes on a larger scale than just the projects studied. It is also difficult to make statements about the achievement of societal impact. After all, the starting point is an explanation of the causes of a positive impact, and OH provides no means of refuting the resulting theory.

Contribution analysis & process tracing

Description

[20] 'Outcome Mapping' is also discussed in the Policy Evaluations Toolbox, but although that method is related to Outcome Harvesting, it lends it self more to the advance planning of a change process (rather than subsequent examination of the changes that have occurred).

Contribution analysis and process tracing are qualitative methods aimed at investigating ('opening') the black box of policy. They can provide insight into the operation of policy and the occurrence of desirable developments. The emphasis is on investigating changes that should occur according to a policy theory, without making statements about causality. The usefulness of such analyses lies in being able to trace to what extent desirable developments can be observed in the theory of change, and exactly where stagnation occurs, without definitive evidence being provided for the operation (contribution) of policy. Both methods are frequently used within case studies. However, there are also differences:

- Contribution analysis (CA) can provide more insight into the specific contribution of policy to an outcome by better understanding and explaining the reasons for certain effects and the role of policy and other factors. This is done by testing practice against policy theory, looking at the extent to which results or interim results occur, examining assumptions, checking implementation, testing other influencing factors and excluding alternative explanations;
- Process tracing (PT) starts by formulating hypotheses about cause and effect from the policy theory, describing for each hypothesis which processes and mechanisms would occur if the hypothesis were correct. It also describes which data and which evidence can be found for this. By means of a combination of quantitative and qualitative data collection methods, the process and the results of the practice (and its operation) are then revealed and the extent to which theory is consistent with practice is tested. Alternative hypotheses are also formulated and can be tested.

The main difference between the two methods is that the testing of hypotheses in process tracing is geared more towards identifying the processes operating in practice that can determine the success or failure of policy in a particular context, while contribution analysis is geared more towards the relative importance of different explanations. A common feature is that both methods offer an approach for seeking answers to the question of how and why effects come about, without making statements about the size of the effects. A detailed policy theory is required for the application of this method, as a structure for plotting the impact of policy.

Relevance to S/T policies

A common assumption when conducting S/T policies is that coordination – and hence governance – is essential for system strengthening or transformation and transition. Contribution analysis and process tracing help to show how the characteristics of the chosen governance processes and policy mix add up to a chain of desirable or undesirable outcomes, which may also include system strengthening and structural changes (in other words, the method covers perspectives B-F).

This effect chain is reviewed step by step on the basis of the policy theory. For example, if 'openness' is an important governance principle, this should have consequences for the diversity of actors involved in writing roadmaps, which in turn can influence the composition of project teams, the variation in solution directions in a project portfolio etc. And if a 'focus on selectively chosen topics' is an important governance principle that is also fulfilled in implementation, this should manifest itself in an increased degree of selectivity and coherence in the innovation paths that are central to efforts to initiate a transition, which should in turn be reflected in, for example, the scope of investments and loans or of experiments that arise.

The methods are potentially interesting for the evaluation of S/T policy, because reviewing causal chains not only answers the question of whether changes occur (which are consistent with the policy theory), but also at what point in the policy theory these changes take place. After all, it is possible that the introduction of S/T policy is associated with adjustments to the strategies of influential actors, but that the changes stagnate when the actors' activities have to translate into a change such as a leap

in sustainable hydrogen production. Knowledge of where the chain between policy and desired outcomes 'breaks' is relevant to both formative and summative evaluations. A limitation of the two methods is that they leave little room for 'surprises', since they are based on a predefined policy theory. In that regard, the OH method is more effective when action mechanisms are still open to discussion.

Qualitative Comparative Analysis (QCA)

Description

Qualitative comparative analysis (QCA) is a method used to find out why a policy measure achieves the intended result under certain circumstances and not under others. The aim of the method is to identify conditions (or combinations of conditions) that are necessary and/or sufficient for the occurrence or non-occurrence of a certain outcome. Cases are compared systematically. This does not necessarily result in a single model; the method is open to multiple (possibly rival) explanations for the relationship between condition sets and outcomes that can then be tested. QCA is particularly useful for understanding *what* potentially influenced an outcome.

Relevance to S/T policies

QCA lends itself to analysis of a complex policy reality. One of the principles of the method is its ability to deal with causal complexity (conditions that only produce a certain outcome in combination with other conditions; A does not necessarily lead linearly to B and there may be causal asymmetry). In the complex policy reality of systems and transitions, QCA can be used to search for the conditions that influence the occurrence of certain effects. This requires the existence of several comparable cases, with sufficient variety in outcome variables in order to investigate which combinations of conditions are associated with a better outcome. In addition, the number of possible relevant (contextual and other) conditions must not be too large, the variety of contextual conditions must be low and the variety of 'causal' conditions must be high.

Case selection within S/T policies is a challenge, particularly for evaluations at meso level (of systems or transitions). At first glance, the method appears suitable for making comparisons between system-strengthening networks, as found in the Technology Pact or as in the Smart Industry Field Labs, but if those networks differ from each other in many respects, it is difficult to meet conditions relating to the variety of contextual conditions. In the case of the missions in the mission-oriented Top Sector and Innovation Policy(MTIB), it is therefore difficult to compare the 25 missions and ask which factors are associated with observed performance in achieving the specified (and possibly very disparate) mission objectives. Instead, QCA is better suited to research into a single mission that involves working on parallel but similar tracks, for example when several regions all focus on pursuing the circular economy in the same way. In this situation, it is possible to compare which conditions (variation in the presence of policy incentives, industry, active citizens etc.) can be related to higher performance in circular economy indicators in those regions. A similar application is the use of QCA to explain differences in the success of the Regional Energy Strategies (RESs) of the 30 energy regions in the Netherlands.

Regression analysis

Description

Regression analysis is a statistical analysis that reveals connections between variables – see Effect Measurement Expert Working Group (2012). It can be used to examine the presence and extent of the connection as well as the type of connection (positive/negative). Essentially, the results of a regression analysis only indicate correlation (controlled for factors that may obscure that correlation)

and do not offer any conclusion about causality. However, causality can be made plausible with various methods, for example by applying panel techniques and using discontinuities in various ways. Sufficient quantitative data is a requirement for regression analysis, which may concern data from monitoring or survey research but often also includes administrative data such as that found in Statistics Netherlands microdata.

Relevance to S/T policies

In the case of S/T policies, regression analyses are best used for interventions that address a particular component of a system without many other factors that are difficult to know or measure affecting that aspect. For example, the development of knowledge on a certain subject (such as sensor technology) can be important for a system or transition, and it is conceivable that there are targeted interventions that could be reliably linked to developments in that knowledge production. It is more difficult to make statements about the extent to which increased knowledge production leads to faster development, production and rollout of a particular innovation, because in the case of S/T policies those outcome aspects often depend on other system factors. To determine whether regression analyses can be used, it is therefore important to verify first whether a policy objective really requires strengthening or transforming a system, or whether a straight causal line can be drawn between policy impulses and policy outcomes.

Experimental/quasi-experimental methods

Description

Experimental and quasi-experimental research methods can be used to establish causal relationships between a change in an independent variable (such as better information about working in technology for a group of young children and changes in the dependent target variable of the inflow into technology education) and a dependent variable. In this way, it is possible to demonstrate the effect of a measure for this group relative to another group. Specifically, an experimental research method means that two comparable groups are required, with one group (the experimental group) being exposed to an intervention and the other group (the control group) not being exposed to it, the latter group being randomly composed; randomised trials are an example of this. In a quasi-experimental method, a group of participants undergoing a specific intervention is compared with a control group that does not undergo that intervention. The assignment of participants to the groups is not random, however, hence the term quasi-experimental. Take, for example, policies that are introduced earlier in some regions than in others. The difference in outcome between the two groups can then be attributed to the intervention or the measure.

Relevance to S/T policies

It is important to experiment in S/T policies. This makes it possible to test partial effects of the policy by means of a small-scale experiment in order to learn about the effects of policy incentives. In addition, the effects of individual instruments can be tested by formulating a clear design in advance – see Effect Measurement Expert Working Group (2012). For the S/T policy as a whole, this method will probably be relevant at most for a subset of the policy mix and at output rather than outcome level. This is because outcomes are formulated too broadly and influenced by various factors and instruments.

Monitoring

Description

In the case of monitoring, data concerning organisations, projects, systems or interventions, for example, are collected, stored, processed and reported periodically on the basis of a fixed pattern. With this method it is possible, for example, to monitor the performance of an organisation and to make adjustments when necessary. Monitoring is not by definition tied to the use of a policy measure or objectives of policy or implementation; it is a precondition for measurement. The indicators used can also provide important contextual information for a policy measure.

Relevance to S/T policies

Continuous and long-term monitoring is a precondition for gaining a picture of policy input, progress, output and outcome. Systematically keeping track of policy input and its outcomes is important in order to visualise the validity of policy theory at different points in time. In transitions, monitoring is necessary to illustrate the dynamics of the transition. During a transition, there is a need for indicators showing the phase a transition has reached and how policy fits in with it and/or needs to be adjusted. In the case of a transition, it is also important to measure the extent to which the ultimate objective is and remains in sight. Finally, the monitoring of intermediate goals (system or transition parameters) increases the reliability of conclusions to be drawn about the impact of policy.

6. Step-by-step plan for preparing and designing evaluation studies

With S/T policies, it is often unclear how the policy (by strengthening or transforming systems) should ultimately contribute to the specified policy objectives. Often it is first necessary to clarify, for example, exactly what the S/T policy entails (coordination structures, a policy mix, both), what the underlying policy theory is, which system is involved and how it is delineated.

Concepts and frameworks have already been introduced in the previous chapters. Together these can provide a basis for a structured review of the essence of S/T policies and the methodological possibilities for evaluating them. In this chapter we summarise this in a step-by-step plan. The policymakers are likely to be best placed to take the first steps, with the role of evaluators gradually increasing towards the later steps.

Step 1: What is the objective of the policy?

S/T policies are aimed at producing and distributing public goods and services, these being the result of an interactive process between different actors, such as businesses, universities and the government. If there is no government intervention, the system will be ineffective or the transition will be too slow. The ultimate objective of the policy must be formulated, along with the path towards it, including the transition pathway, the change in the operation of the system (and the functions) and the time by which certain policy objectives must be achieved.

Step 2: Which failures have led to government intervention?

Market, system and transition failures must be analysed to identify the rationale for government intervention in concrete terms, i.e. what specific problems the policy should address (in order to influence an ultimate societal objective effectively and efficiently) and what type of policy that requires. The extent to which the government can resolve or reduce the failure through policy, or whether there is a reasonable likelihood of government failure, must also be assessed.

Step 3: How does the system work or how is the transition designed?

An evaluation of S/T policies starts by delineating the system or transition and the evaluation period. An important question is whether a top-down approach can be adopted in which the instruments relating to a budget item on a certain subject play the dominant role (as in a policy review), or whether a bottom-up approach is adopted in which the policy is delineated by examining all policy instruments belonging to a system or transition.

The intervention logic follows from this. This logic consists of exploring the characteristics that determine what the policy aims to change in the functioning of a system or bringing about a transition and how those intended objectives are achieved. In order to structure this process, it is necessary to answer eight questions:

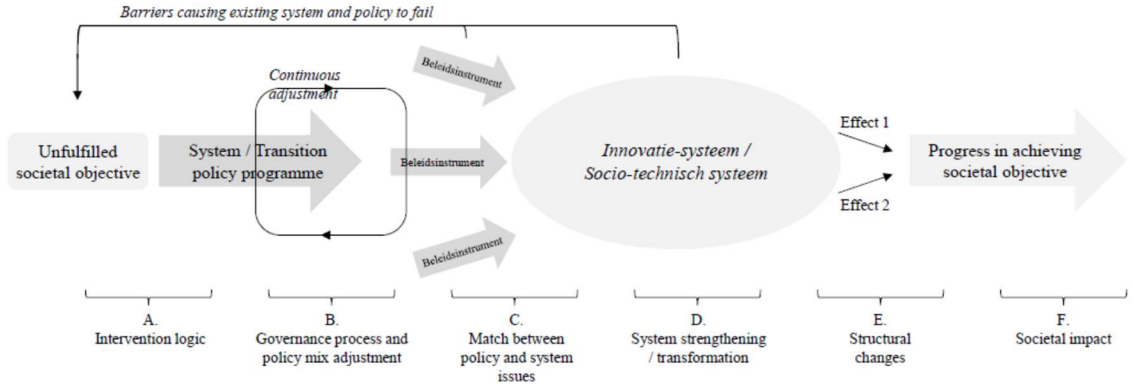
1. Definition. What is the definition of the system or the transition?
2. Objective. What are the specific problems that the policy is seeking to remedy, and what are the reasons for them?
3. Policy. What policy activities and instruments in the S/T policy can be used, who will implement them and what is the budget?
4. Contribution. How will these efforts and instruments contribute to the intended objectives and by when?
5. Coherence. What coherence is there between the efforts and instruments?
6. Monitoring. How do any monitoring activities and indicators relate to the operating mechanisms and objectives?

7. Government failure. What are the reasons and indications for the non-operation of the policy?
8. Nature of the policy. To what extent is the policy dynamic or adaptive?

Step 4: Which part of the system or transition needs to be evaluated?

A system or transition is not usually evaluated in its entirety, because a system evaluation is about the extent to which different functions in the system are effective and in the case of a transition the objectives lie in the distant future. There are six perspectives for evaluation (Figure 6.1). Which perspective is most relevant will depend among other things on where an S/T programme stands in the policy cycle. When the programme has only just been launched, it may already be possible to identify which instruments it will initially want to direct, but that coordination or coherence can only be evaluated once a period has elapsed during which that policy mix was in force. Improving on transforming systems requires a lead time. The legitimacy of the intervention (is there a clear issue justifying the government intervention?) and the governance processes can already be evaluated when a new programme is deployed, but can also be relevant if that programme continues for longer.

Figure 6.1. Ordering of the six perspectives for the evaluation of S/T policies.



Step 5: Formulate hypotheses about what the policy should achieve and link these to the policy theory

Hypotheses must be formulated with regard to what the policy should achieve. Hypotheses ensure that the evaluation can focus on an assumed mechanism and an intended objective.

- Core hypothesis: formulate the main objective of the system or transition.
- Hypothesis concerning direct effects: formulate a hypothesis that states in concrete terms what effect is expected from the part of the policy to be evaluated.
- Hypothesis concerning indirect effects: formulate a hypothesis about spillover effects or higher-order effects of the policy.
- Hypothesis concerning societal effects: formulate a hypothesis about the impact of the policy.

Hypotheses can be conceived for each of the six perspectives, but they have a somewhat different character. That means the hypotheses on the left of Figure 6.1, for example, are mainly about whether the coordination and the instrument mix are appropriately organised and to what extent the system or transition will have the desired direction and progress at the desired pace, while the right-hand side will address the effects of policy on systems, structures and ultimate objectives.

The hypotheses can usually be linked to a mechanism in the policy theory and in that regard also constitute an assessment of whether the mechanisms assumed by the policy theory also occur in practice. In the case of hypotheses relating to the left-hand side of Figure 6.1, it is possible that very little can be said about the higher-order goals. The analysis of output in terms of realised governance processes and their influence on the instrument mix is probably still achievable in many cases, but the impact of governance processes on first-order and higher-order objectives is empirically more difficult or impossible to determine. On the basis of the policy theory, there may also be additional hypotheses, in which case it is necessary to return to step 5.

Step 6: Identify frameworks, associated indicators and methods to assess performances, preferably on the highest possible steps of the effect ladder

The next step is to determine the evaluation method. Conventional effect measurements as discussed in the report of the Effect Measurement Expert Working Group (2012) often involve selecting or constructing a control group, or another kind of counterfactual (e.g. current performance compared to past performance). Ideally, such techniques will also be used in the evaluation of S/T policy, especially for the output, outcome and impact objectives.

Where that is not possible, it will be necessary to rely on alternative assessment frameworks. To this end, on the basis of a formative evaluation approach, it is possible to consult stakeholders who have an idea of how the policy should function. An alternative route for the creation of assessment frameworks is to select and possibly combine theoretical frameworks. Some examples of this were discussed in Chapter 4. The advantage of this is that the frameworks are relatively more tried and tested than the ideas of stakeholders, particularly if they ensue from a rich research tradition with empirically rigorous methods. A disadvantage is that the frameworks are not always a perfect match for a specific case. For that reason, it is important to state explicitly up to what level a theoretical framework is useful for the evaluation of a particular policy aspect or hypothesis. It may be necessary to integrate multiple frameworks, or to translate them by other means into the context in question, in accordance with the reflexive evaluation method. Here too it may be possible to call on stakeholders who understand the reasons for framing Dutch policy in a particular way.

Step 7: Measurement – Perform the individual sub-analyses

When all relevant conceptual frameworks have been selected and operationalised, the actual measurement can take place. Among others, the quantitative and qualitative methods discussed in Chapter 5 can be used for this purpose. There are a limited number of methods that can be used to measure the size of effects, whereas there are various alternatives for research into operating mechanisms and links.

Which method is most suitable for an evaluation will depend among other things on the precise evaluation question and data availability. Table 6.1 below summarises the relationships between the six evaluation perspectives and the methods that can be used. A filled cell means that the method will probably be useful (e.g. for exploration, assessment) in order to make statements about that perspective. When selecting a method, it is essential to ask what is already known about the operation of the policy, as some options are mainly useful for revealing and understanding operating mechanisms (which in S/T policies can rarely be simply inferred from policy documents) and other options are more suitable for establishing empirically whether and when effects occur, how they occur and how large they are.

Table 6.1. Overview of research methods and the perspectives for which they appear most appropriate.

	Intervention logic	Governance and policy mix	Match between policy and issues	System strengthening / transformation	Structural changes	Societal impact
Effectiveness						
Systematic reviews & meta-analysis	Systematic reviews involve identifying frameworks and principles against which to test policy implementation. A meta-analysis is a numerical analysis.					
Reflexive evaluation	Explaining and (possibly jointly) interpreting mechanisms, starting from policy theory and implementation					
Case studies			Explaining mechanisms & determining outcomes			
Outcome harvesting			Explaining mechanisms, starting from outcomes (how does that outcome arise; what is the role of policy?)			
Contribution analysis & process tracing		Determining causality, starting from policy theory (gathering indications for the occurrence of a chain of outcomes)				
Qualitative Comparative Analysis (QCA)			Determining causality, starting from outcomes (what combination of factors determines variation in success?)			
Regression analysis				Measuring the size of effect, correcting for control variables		
Experimental/quasi-experimental methods				Determining causality & Measuring size of effect		
Efficiency						
Monitoring		Keeping track of developments, in predefined indicators				
Productivity examination (DEA, SFA, benchmark)				Comparing costs and/or performances (or the relationship between them)		
Simulations*					Estimating outcomes of complex interactions (including policy interactions)	
Societal cost-benefit analysis (SCBA)						SCBA

*Simulations can also indicate effectiveness, but they are included here only once for the sake of clarity

Table 6.1 suggests that there are few methodological tools for examining the efficiency and effectiveness of policy objectives focused on governance and adjustment of the policy mix, whereas these aspects are crucial for experimental S/T policies based on the notion that "it is important to do something about societal objective X, even if we do not yet know what works".[21] This is because the assessment of governance has more to do with evaluation criteria such as coherence/consistency, high-quality implementation and support for policy – and therefore does not directly describe effectiveness or efficiency. It is true, of course, that the proper organisation of policy can be a prelude to generating effects; hence the place of perspective B in the broader evaluation framework.

Step 8: Policy choices

In S/T policy, the way in which policy is designed and implemented is important. If the picture of the future is uncertain, three possible forms of implementation should be visible.

The first is the option to wait before implementing the policy if *ex ante* costs and benefits indicate that higher benefits can be achieved with the same effort in the future. Considerations that must be visible

[21] This starting point (we do not know if our policy is working, but the problem is urgent and doing nothing will make the problem worse) is referred to as input legitimacy. This is different from output legitimacy; this states that policy is legitimate if it can be demonstrated to be effective and efficient.

consist of recorded decisions to deploy a policy from a given point in time, a consideration of the intensity and scope of the policy and an estimate of the costs and benefits.

The second form concerns dividing up policy. A factor here is the extent to which there is an efficient choice of various options with explicit attention devoted to costs, the option to halt the policy and the consideration of uncertainty and risks.

The final form concerns experimentation. Uncertainty often arises due to knowledge gaps. In such cases, the option to close these gaps through experimentation should be visible. It is appropriate to include a judgement as to whether the experiments are the right ones and the extent to which they have been carried out efficiently, as well an opinion on the lessons learned for policy.

Step 9: Combine the insights from the complementary partial analyses

Given the relative advantages and disadvantages of different types of analysis, this final step prescribes a synthesis in which the sub-outcomes (possibly from different evaluations) are compared. Insight into perceived strengths and weaknesses in the governance structure can be valuable in itself, but it is even more useful if there are also indications of the effects of the policy – and vice versa. When combining and reporting results from sub-analyses, it is essential to stay focused on the kinds of statements that can be made on the basis of the methods used. It is particularly important to avoid talking about overall policy effects when these have not been measured reliably.

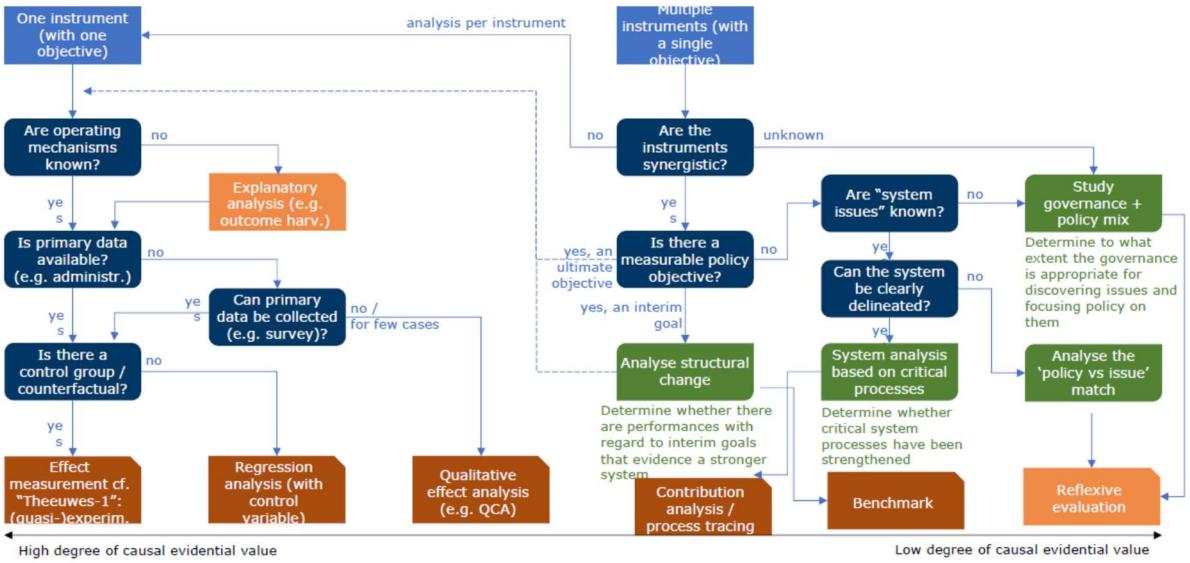
7. Decision aid for selecting methods

To assist evaluators in selecting methods, this chapter presents a decision aid in the form of a flowchart (Figure 7.1). This is not a model in which an answer rolls out mechanically, but a chart that can be used to systematically assess which analytical frameworks and methods best suit the characteristics of a concrete case of S/T policy. It therefore also forces us to think about those characteristics that are seldom evident in S/T policy and were discussed in the earlier chapters.

By going through this chart from top to bottom, evaluators can see the empirical methods that appear most suitable for the evaluation of a particular case. These methods are broadly ranked according to the level of causal evidential value they can offer, in accordance with the 'effect ladder'; see Section 4.1. On the left of the figure are the methods for which the causal evidential value is the highest, with the assumption that there is also a preference for these in effect measurements. The report of the Effect Measurement Expert Working Group has already described the specific techniques that can be considered when using experimental or quasi-experimental methods. Although the axis is not completely ordinal, the causal evidential value of methods decreases the further they are to the right in the figure.

The methods to be considered are shown in the dark orange blocks. As discussed earlier, the method on the farthest right, reflexive evaluation, does not lend itself to statements about causal effects. This method block therefore has the same lighter colour as the 'explanatory analysis' block. The green blocks refer to the analysis frameworks that appear depending on the characteristics of the policy to be evaluated and the available information on it. They are linked to the perspectives from the evaluation framework in Chapter 4. Table 6.1 has already shown that there is a connection between those perspectives (and associated analysis frameworks) and suitable empirical methods.

Figure 7.1: Decision aid for selecting empirical methods for the evaluation of S/T policies



7.1. Decision aid for policy with one instrument and one objective

For evaluation studies focused on policy relying on only one instrument but with one policy objective, there are relatively good options for applying the most desirable evaluation standard, the experimental or quasi-experimental method. That can be used when three conditions are met:

- there must be a clear picture of the operating mechanisms of the policy, so it is clear which indicators can be assessed, with whom, at what time (depending on the 'incubation time' of policy incentives and outcomes) etc.;
- data must be available on, for example, the incentives provided and performances, preferably primary objective data as available in administrative reports or from Statistics Netherlands; and
- there must be a counterfactual (relevant basis for comparison) such as a control group, or at least it must be possible to construct one (e.g. using econometric models such as propensity score matching).

If it is not possible to fulfil the latter condition, evaluators can resort to regression analyses to determine whether the treatment group shows a statistically significant connection between policy incentives and performances. Then, with the aid of control variables, a correction must be applied as far as possible for factors that also affect the analysed performance. However, the causal evidential value will be lower than if a valid comparison can also be made with similar parties that have not been treated.

If in addition no relevant primary data are available in administrative sources (which may still need to be interlinked), that problem must first be remedied. The question is then whether such data can be collected. If primary data cannot be collected, or can only be collected for a handful of respondents, it is possible to resort to structured forms of quantitative effect analysis.

If no clear picture whatsoever can be obtained of the mechanisms by which policy is ultimately deemed to contribute to policy objectives, this must first be ascertained. For this it is possible to rely on explanatory analyses such as outcome harvesting and case studies. By using such methods to assess how the desired interim and final outcomes are achieved, it is possible to continue on the flowchart and thereby determine which method can be used to also carry out a causal assessment of whether such outcomes are attributable to policy.

7.2. Decision aid for policy with multiple instruments and one objective

The first question is whether the instruments in an S/T policy programme are synergistic This is the case if those multiple instruments influence the same system or the same transition, but it is not possible to distil the individual influence for each instrument. The flowchart offers three options:

- The instruments are not synergistic or are only synergistic to a limited extent. In that case researchers can go through the left-hand side of the figure for each instrument;
- It is not known whether the instruments are synergistic. This situation arises when the precise operation of the S/T policy is not sufficiently visible. This is the case, for example, if policy focuses primarily on coordination and streamlining all kinds of new and existing instruments and the related content. Investigating whether such policy makes a difference requires a precise understanding of the organisation of the policy and the relationship with existing instruments. The counterfactual is not that all those instruments would not exist, only that they would not be controlled from the S/T programme. The situation can be clarified by means of analysis frameworks associated with **perspective B** in the evaluation framework. **'Governance and policy mix'**. As described in Section 3.4, this is about assessing the governance, and determining the extent to which it is suitable for focusing policy on the most urgent issues in a system or transition. Normative evaluative statements require principles to be identified which the governance must fulfil. The 'reflexive evaluation' method therefore appears to be perfectly relevant here, since participation by policymakers and implementers is essential in determining such principles;
- The third possibility is the assumption that policy is synergistic. This leads to the question of whether a measurable policy objective is also linked to it.

The question with regard to a measurable objective concerns the availability of an indicator that helps to verify whether there are developments in the variable that the policy seeks to influence. Here too there are three options, including two variants of 'yes, there is an objective':

- There is a measurable final objective; the variable that ultimately matters can be measured directly. If this is the case, an attempt can be made to view the overall synergistic policy mix as a single combined policy incentive, and to develop an evaluation design for it in accordance with the left-hand side of the flowchart. Ideally, the main points of the operating mechanisms will be known, data will be available and a counterfactual can be found such as the situation before and after the introduction of the S/T policy mix. The line towards the left-hand side of the flowchart is dotted, because it is not certain that a suitable method can ultimately be selected. For example, if the different components of the S/T policy mix have been implemented very gradually, or if there are exogenous factors that have clearly also had a major influence on the policy objective, it becomes difficult to formulate a solid causal analysis;
- In some cases there will also be one or more intermediate goals that first have to be attained in order to achieve the intended impact. Hence these are not temporal intermediate goals, but goals that in the theory of change precede the ultimate policy objective. In order to evaluate policy effects on that intermediate goal, it is possible to use **perspective E** from the evaluation framework, to analyse '**structural changes**'. This concerns changes (in this case) in the economic structure that is important for achieving the ultimate policy objective. Various methods can be considered:
  - Once again it is sensible to determine first whether the structural change can be evaluated through the left-hand side of the flowchart. In the details of the MTIB case in Chapter 8, for example, we see that there appear to be possibilities for determining whether individual businesses (according to the WBSO records) have indeed focused their own innovation projects more on the themes boosted by the MTIB coordination and policy mix. This could be compared with innovation activities for similar economic activities or technologies that fall just outside the scope of the KIAs and MMIPs.
  - If analyses with a high causal burden of proof are not achievable, it is possible to resort to methods such as contribution analysis and process tracing. These provide a basis for systematically revealing which changes have taken place in all parts of the effect chain through to the intermediate goal to be investigated. If there are positive developments on all these points, that does not yet provide causal evidence, but such a reconstruction does provide useful information about whether desirable developments are taking place at all. If that is not the case, it may be an indication of a lack of policy effectiveness;
  - An alternative method is to benchmark measurable structural changes, for example by looking at other regions or countries. When designing and interpreting such a benchmark, careful account must be taken of the extent of comparability: is there no policy elsewhere (and, in terms of policy effectiveness in the Netherlands, is it therefore a bad sign if the structural change proceeds just as quickly there), or is there the same type of policy (and is it only concerning if the Netherlands lags far behind)?

A third possibility in the flowchart is that the S/T policy does not have a measurable ultimate objective. This is typical in the case of system policy that seeks to promote the functioning of an innovation system as a whole, unlike in the case of transition policy aimed at the concrete and measurable societal objective with regard to carbon reduction, for example. If the S/T policy is aimed at strengthening systems, as was the case of the original Top Sector policy, the next question is whether the main issues in those systems are known. The flowchart can be followed in two ways:

- No, the system issues are not known. In that case the only option is to examine the quality of **'Governance and the policy mix'**, according to **perspective B**. Attention can also be devoted to the question of whether the governance and the policy are organised in such a way that it is possible in any event to learn what the system issues are. In the fields of both innovation policy and climate policy, there are policy strategies that are not based on a known set of issues, but which try to ascertain by means of experimentation why certain desirable changes do not occur. An evaluation could thus focus on developing the analysis framework by formulating/identifying appropriate principles, which are important when learning about issues. Here it is possible to adopt the reflexive evaluation approach, so as to guarantee a degree of 'reflexivity' (reflection on policy objectives, frameworks, organisations etc.);
- Yes, the system issues are known. It is clear that specific processes such as knowledge development or training, for example, are not well geared to the occurrence of desirable economic or societal changes. In order to assess whether policy has brought about a change, the system must be clearly delineated; in which sector/technology area/region should the weak processes now be strengthened? Here again there are two possibilities:
  - The system can be clearly delineated, in a such a way that data can be collected and analysed for that system. In that case, an evaluation will need to demonstrate whether the system has begun to function better after a period of time. For this it is possible to use the analysis frameworks associated with **perspective D, 'System strengthening/transformation'**. An example is the TIS framework with seven key processes, as also used in the evaluation of the Top Sector approach. In terms of empirical methods, the main options are the same as in the analysis of structural change;
  - Finally, it is also conceivable that the system cannot be clearly delineated. This will be the case if the system is described in very general terms, or if the boundaries cannot be readily turned into analysis variables. A 'regional innovation system', for instance, does not always fall within the administrative regional boundaries, and often connects a set of actors that cannot easily be described using sector codes or technology fields. It is then impossible to verify whether the relatively intangible system is functioning better. A remaining option, however, is to determine whether the policy is at least tackling the known issues. For this, see **perspective C, 'Match between policy and issue'**. Research based on the analysis frameworks provided in Section 4.4 says nothing about the size of causal effect, but it is likely that policy will be more effective if it tackles urgent issues than if it focuses *de facto* principally on processes that do not cause any problems.[22] In this type of analysis, the 'reflexive evaluation' method should be considered, since this is a way of checking, with policymakers and implementers among others, how the policy impulses are working in practice. As stated on previous occasions, this is often unclear in S/T programmes that mainly consist of agendas or coordination structures rather than subsidies or rules.

[22] The analysis proposed here assumes that the system issues are already known. In practice, there may well be evidence that the issue concerns, for example, a process such as 'knowledge dissemination' or 'market creation', but that there is no clear indication of the cause of this issue. In that case, evaluators or other researchers can first refine the problem diagnosis, for example by assessing whether relevant actors have insufficient knowledge to show a desired behaviour, or whether they are *unable* or *unwilling* to show this behaviour. The sharper the problem analysis, the better the evaluators can investigate whether the policy is well targeted.

8. Initial lessons based on exploratory cases

The usefulness of the frameworks and methods discussed in this report, and specifically the decision aid in Chapter 7, is best demonstrated by a specific application of it. The Ministry of Economic Affairs and Climate Policy has therefore produced an inventory of cases that could potentially benefit from the content of this report. On this basis, four cases were selected that differed in the degree and manner in which they exemplified system and/or transition policies. The variation on this point allows an exploration of the opportunities and limitations of the tools provided. The four selected cases are:

- The Mission-oriented Top Sector and Innovation Policy;
- The CO_2 reduction policy / Climate policy;
- The Dutch Digitisation Strategy;
- The Technology Pact.

The detailed case studies, which have been drawn up in collaboration with the responsible officers in the Ministry Economic Affairs and Climate Policy, are available on request. For this report they were instrumental in assessing the factors that evaluators can use as their starting point when formulating and implementing evaluations. In this chapter we report first on how the decision aid for each of the cases helps in delineating and implementing evaluation designs. We then reflect on a number of lessons learned across the four cases.

8.1. Application of decision aid to each case

The Mission-oriented Top Sector and Innovation Policy

Description

The Mission-oriented Top Sector and Innovation Policy (MTIB) builds on the Top Sector policy launched in 2012 by the ministries of Economic Affairs and Climate Policy and Education, Culture and Science. The heart of that policy comprises the nine Top Sectors; techno-economic fields characterised by a strong research and innovation profile. Each Top Sector has a Top Team with representatives from science, business (including an SME representative) and relevant specialist departments. These Top Teams are tasked with increasing the opportunities for innovation in their field. For this they can rely on the Top Consortia for Knowledge and Innovation (TKIs), which are largely financed by the Ministry of Economic Affairs and Climate Policy and serve as secretariats. The TKIs deal with matters such as network activities and roadmap development, in order to create ecosystems and reveal what the innovation ambitions are and what obstacles stand in their way. This also gives rise to the Knowledge and Innovation Agendas (KIAs) of the Top Sectors. The Top Teams use the KIAs and underlying signals to influence matters such as research funding, regulations, export promotion and curriculum development. In order to align research programming in the Dutch innovation system with the KIAs, Knowledge and Innovation Contracts or Covenants (KICs) are signed every four years by the Top Sectors, the relevant ministries and knowledge partners such as the TO2 institutes. In addition, there are some schemes specifically for the Top Sectors and the Top Sector policy. The largest of these is the PPP allowance, the instrument that enables TKIs to pursue public-private R&D projects, as they receive a 30% allowance on top of private contributions to such projects. A precondition, however, is that both the allowance-generating 'basic projects' and the 'deployment projects' (what the allowance will be used for) fit within the KIAs. In order to ensure that SME projects are better aligned with the KIAs, there is a separate scheme, the Regional and Top Sector Innovation Scheme for SMEs (MIT). Activities funded in this way this include knowledge vouchers, feasibility projects, R&D cooperation and TKI innovation brokers.

The ultimate objective of the original Top Sector approach was to increase private contributions to public-private research, and ultimately to strengthen the Netherlands' competitiveness. Over the years, however, the emphasis has gradually shifted to the possibilities of focusing the collective innovation efforts more strongly on societal challenges. One of the factors here was that European R&D policy focused on themes such as 'grand societal challenges' and 'societal missions', and that some specialist departments wanted a greater say on the direction of innovation in the Top Sectors relevant to them. In the development of the KIAs for 2018-2021, the Top Sectors were asked in mid-2017 to compile these on the basis of societal challenges and key technologies.[23] In mid-2018 the Ministry of Economic Affairs and Climate Policy announced that these would be the leading themes for the new generation of Top Sector policy.[24] From 2019 this movement was continued with the switch to the Mission-oriented Top Sector and Innovation Policy.[25] This is based on the 25 missions associated with four overarching mission themes (see Appendix), combined with a policy track to promote key technologies.

In terms of both governance and policy instruments, the MTIB is firmly anchored in the Top Sector policy. The switch to the MTIB implies that changes must be made to existing structures, such as the establishment of 'Mission teams'. In the case of missions that rely heavily on a single Top Sector, these are in addition to the Top Teams, whereas, for example, in the mission relating to the Circular Economy (which can access innovations from multiple Top Sectors), they are a relatively new body. The establishment of the MTIB was initiated in 2020, with the speed of developments varying among the various missions.

The figure below outlines the policy structure of the MTIB:

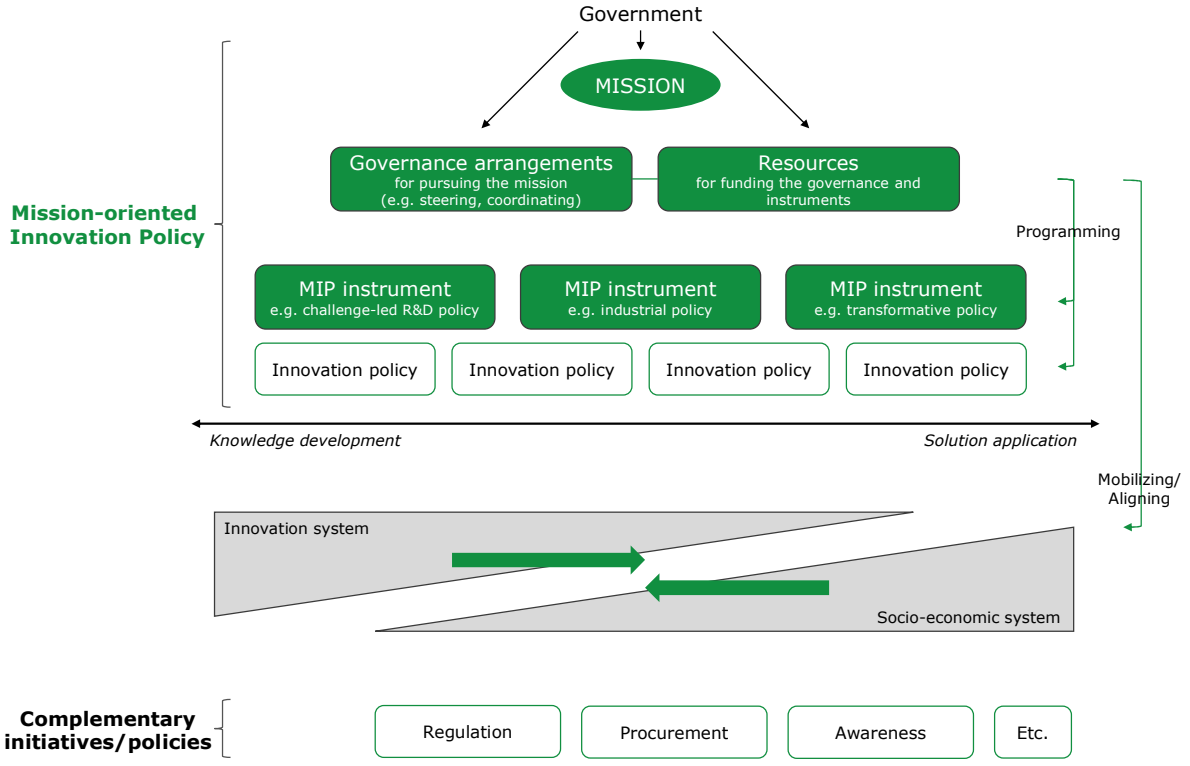


Figure 8.1: Structure of the MTIB, in outline (source: 'Post-commencement assessment MTIB'. Janssen, 2020)

[23] Knowledge and Innovation Agenda 2018-2021: Societal challenges and key technologies.
[24] Ministry of Economic Affairs and Climate Policy (13-07-2018). Letter to Parliament on innovation policy and innovation promotion: towards mission-oriented innovation policy with impact.
[25] Ministry of Economic Affairs and Climate Policy (26-04-2019). Letter to Parliament on Missions for the Top Sector and Innovation Policy.

The essence of the MTIB is that, as in the Top Sector approach, governance structures have been established to schedule the innovation efforts. For this purpose, those structures primarily use existing instruments of the Ministry of Economic Affairs and Climate Policy and other government bodies. The ambition is to have better coordination between investments in research and innovation on the one hand, and efforts to remedy societal problems on the other. Connecting the two associated systems should provide a better perspective on innovative solutions that benefit mission objectives.

Considerations based on the decision aid

Below we go step by step through the questions in the flowchart. The starting point is that we are dealing here with a policy in which multiple instruments jointly pursue one goal.

Are the instruments synergistic?

The synergies intended in the MTIB relate to the strengthening of system factors that are precondition for the success of specific innovation paths (namely those which the coordinating bodies believe can contribute to the achievement of mission objectives). It is possible to strengthen only knowledge development or market formation for such a path, but, particularly in the case of 'wicked' societal challenges and transitions, the principle is that new movements will only emerge if many lights are switched to green at the same time. This is also the central idea in the pursuit of a continuum in the provision of policy support; policy incentives must complement each other in such a way as to eliminate obstacles on innovation paths. The clearance of diffusion obstacles, such as obstructive regulation, can create a demand pull for innovation activities that are still in an early development stage. Hence it is not easy to compare projects from one MTIB-related scheme with thematically similar innovation projects that have not received any support from that scheme. After all, the latter innovation projects could also have benefited from the more favourable conditions for innovation on that theme.

Is there a measurable policy objective?

The MTIB as such does not have a concrete ultimate objective, as it mainly creates conditions enabling a diverse set of public and private partners to collaborate on achieving the 25 missions. Many of those missions each have their own concrete ultimate objective, however, and even a specific time by which it must be achieved. Examples are "A carbon-free built environment by 2050" and "In 2030, care will be provided 50 percent more (or more often) in the recipient's own living environment, rather than in care institutions". In the case of other missions, the objective is formulated fairly ambiguously, such as "in 2035 the Netherlands will have the navy of the future, one that is able to respond flexibly to unpredictable developments".

In any event, it is also advisable to investigate for each individual societal mission whether it is possible to follow the left-hand side of the decision aid flowchart. For the MTIB itself, which supports the development and adoption of innovations that can benefit the mission objectives, this option is unavailable due to the lack of a concrete measurable objective. Hence two options remain. It can be stated that there is no policy objective (as detailed further below), but it is also possible to assert that the MTIB is aimed at pursuing *"a shift of people and resources to areas where they have most impact (= thematic convergence in innovation efforts)'.* Such convergence is a uniform interim goal as a step towards achieving mission-specific objectives. We discuss which methods lend themselves to the interim goal of evaluating thematic convergence at the end of this case.

Are the system issues known?

The question that follows is whether system issues are known. The answer to this again depends on precisely how the MTIB, as the overarching framework for prioritising the missions and organising the structures for the pursuit of those missions, is interpreted. Interpreted broadly, the MTIB relates to the entire range of instruments for achieving mission objectives (including the policies of departments

that have committed to those objectives) and must therefore be accounted for on that basis. A narrow interpretation of the MTIB is that it has put specialist departments in a position to formulate and pursue missions, but that the MTIB 'itself' then only concerns the provision of potentially suitable solutions. In that perspective it is only part of the Dutch mission policy, with the latter also comprising policy efforts which are not directly associated with the Top Sector and Innovation Policy of the Ministry of Economic Affairs and Climate Policy, but which are, for example, associated with legislation and diffusion policy from the specialist departments. The fact that the provisional monitoring framework for the MTIB focuses on techno-economic indicators suggests that the narrow interpretation applies there.

If we use the narrow interpretation of the MTIB, the main issue is the lack of coherence in research and innovation efforts in the Netherlands. The literature on rationales for (possibly transformative) innovation policy also refers to this as 'directionality failure'. The MTIB, the Dutch variant of mission-oriented innovation policy, is an attempt to remedy this problem by tackling urgent societal challenges as pathways for combining innovation investments and capacities and eliciting multidisciplinary cross-sectoral innovations.[26] The structure that has been established serves both to specify the pathways (including the translation from problem to solution) and to continuously adjust policy impulses. The latter is particularly relevant at a time when solutions are continually encountering different problems in their development, which is characteristic of innovations that are intended to bring about transitions.

Although it is therefore possible to identify the failure underlying the MTIB in an abstract sense, it is not easy to be more concrete in terms of system issues. There are probably major differences between the various missions in terms of which processes in particular (e.g. knowledge development, training human capital, eliminating legal resistance) are obstructing the development and application of innovative solutions. In any case, the letters to parliament concerning the MTIB do not mention any general recurring problems that impede progress. As a result, the flowchart leads to the recommendation first of all to conduct a better examination of the governance and policy mix. For the sake of completeness, we reiterate that this reasoning is based on the requirement of evaluating the MTIB as a whole. If the system issues in individual missions are known, the next question is whether it is possible to clearly delineate the associated systems. This seems realistic for concrete missions involving, for example, the Dutch navy or the sustainability of the built environment. In those cases, the decision aid leads to 'System analysis based on critical processes', for which analytical frameworks can be used such as that for Technological Innovation Systems (see Section 4.4) or the mission-adapted variant of Mission-oriented Innovation Systems.[27] We will not consider these analyses here, however, partly because the evaluation of the Top Sector approach is already an example of this; see Box 2 in Section 3.5. That evaluation implicitly used contribution analysis, since qualitative and quantitative indications were collected to determine the extent to which TIS processes were strengthened.

→ *'Study governance and policy mix'*

When examining the governance of the MTIB, the question is whether the coordination structures introduced (theme teams, mission teams, TKIs etc.) are suitable for remedying imperfections in the mix of policy and other incentives, thus facilitating the development and rollout of innovations that can achieve ambitious mission objectives.

[26] Mazzucato, M. (2018). Mission-oriented innovation policies: challenges and opportunities. *Industrial and Corporate Change*, *27*(5), 803-815.

[27] Hekkert, M. P., Janssen, M. J., Wesseling, J. H., & Negro, S. O. (2020). Mission-oriented innovation systems. *Environmental Innovation and Societal Transitions*, 34, 76-79.

The added value of examining governance aspects here is the provision of somewhat greater insight into the quality of policy implementation and the obstacles encountered. Such results are inconclusive about the effect of the policy, but do help to understand what is going right and wrong in management and control – and what could be improved. In cases such as the MTIB, moreover, research into coordination and management forms a basis for a subsequent analysis of how that management is conducted (see below). After all, if the emphasis of the policy is unclear, it is also impossible to assess which movements it will trigger.

Investigating governance is not so much a question of finding the right empirical methods (it often amounts to interviews and desk research), but above all asking the right questions. As explained in Section 4.4, evaluation from the perspective of governance (and the subsequent influencing of the policy mix) benefits from an 'assessment framework' setting out which principles the policy must comply with. The crux in this case is therefore to identify policy principles that can serve as a basis for statements about the extent to which the MTIB is well designed in terms of governance.

There are two routes by which these principles can be clarified: by co-production, entering into dialogue with the field (including the designing and implementing policymakers) and by consulting relevant scientific literature. The **reflexive evaluation** method combines those two routes, since literature analysis can form the basis for developing a theoretical framework (with principles) that can be used to assess policy. In the case of mission policy such as the MTIB, it seems sensible to at least appeal to the academic debate surrounding 'mission-oriented innovation policies' that has gained momentum in recent years.[28] On this basis, the evaluation of governance processes should focus on elements that are of great importance if actors from the innovation system and from social sectors are meant to provide a joint stimulus for the development and application of innovative solutions to the problems on which the MTIB's mission objectives are based. Below are a number of questions that can arise. Note that some questions are based on the presumed need for a certain governance aspect (e.g. there must be checks and balances), while other questions cannot at present go further than revealing the advantages and disadvantages of a particular governance aspect (e.g. is it desirable or undesirable if governments focus on existing rather than new policy on missions?).

Sample questions (indicators) for research into the governance of the MTIB.

- "Has the MTIB led to the emergence of governance structures that are appropriate for coordinating innovation and introduction initiatives with each other?" Appropriate refers to principles as established on the basis of the theory and/or in interaction with the field (see above).
    - Does the structure contain elements that are equipped to collect information with regard to emerging innovation opportunities and mission -related problems?
    - Does the structure contain elements that are able to combine this information and take decisions on the course to be followed? How selective are the programmes (e.g. MMIPs) that arise? And how do they behave?
    - Does the structure contain well-functioning checks and balances to prevent overly strong assertion of interests?
    - Are there indications that the parties affiliated with the MTIB will develop such strong joint plans that they believe they will garner additional support for them (e.g. from the Growth Fund)?
- Are there indications that the MTIB is catching on in terms of mobilising partners? This concerns the actual buy-in by line ministries, regional authorities, knowledge partners etc. By signing the KIC, many parties have already committed resources to the mission.
    - To what extent have these parties listed existing budgets that are used relatively separately from the MTIB and MMIPs? Are there indications that the various partners are actually willing to pool their budgets to achieve a comprehensive and consistent MIP approach (e.g. as in jointly funded policy schemes such as the MOOI?). Do the annual budgets rise?
    - Do the line ministries actively follow developments in the field of MMIPs, as part of a formalised policy cycle?
- "Is the 'comprehensive' MTIB strategy (which also includes line ministries' initiatives) based on a policy mix and funding streams suitable for supporting the entire spectrum of innovation development and deployment?".

[28] For a concise review, see: Janssen, M., Torrens, J., Wesseling, J., & Wanzenböck, I. (2021). The promises and premises of mission-oriented innovation policy: A reflection and ways forward. *Science and Public Policy*, 48 (3), 438-444.

- Is the programming of subsidies for scientific research geared to the knowledge needed in R&D (including PPP) projects related to certain innovations? Is the innovation policy for the most central TRLs (Technology Readiness Levels) synchronised with initiatives to support the introduction?
- To what extent is it problematic if discrepancies exist; does it hinder continuity, and/or does it lead to a critical reassessment of the innovation processes that are being pursued?
- Are there indications (e.g. from data on programming and project portfolios) that the policy mix also allows adaptation of low-TRL instruments in response to issues encountered in high-TRL innovation activities?

As indicated above, it is also possible to take a different turn in the flowchart, based on the interpretation that the MTIB does indeed have a clear intermediate goal with its focus on 'thematic convergence in innovation efforts'. This would lead to another part of the decision aid:

→ *Analyse structural change*

If the aim is to apply more focus and acceleration in innovation paths with potential for missions, effect measurement should indicate whether there is indeed more momentum in innovations in specific directions. The key question is then not whether more attention will be paid to those innovation directions *within* the mobilised schemes, but whether all those policy incentives and innovation projects will initiate so much ("system strengthening") that individual organisations from *outside* the direct target population will also join and invest in those innovation directions.

The most promising means of observing such 'crowding in' is to examine the extent to which innovation projects in the WBSO scheme conform to those directions. Since the WBSO is a relatively generic innovation scheme, without substantive direction, the effect of the MTIB could be demonstrated by an observed increase in the willingness of businesses to perform R&D in products and processes that fit in with the policy-supported innovation paths. If such an increase is not perceptible, it is unlikely that the MTIB has had much impact. If the increase is perceptible, corrections should of course be made as far as possible for any autonomous trend, for example by making a comparison with the development of WBSO projects on themes that also concern the missions' societal challenges, but which have not been selected in the agendas and programmes (KIAs and MMIPs) to achieve the mission objectives. In line with the above measurement, i.e. an analysis of increases in private and applied innovation activity in MTIB innovation directions, it is also possible to examine the economic performance of the parties that have WBSO projects aligned with the prioritised innovation directions.

In short, it therefore seems possible to conduct the research into structural changes using empirical methods on the left of the decision aid (quasi-experimental or by means of regressions). The operating mechanisms are sufficiently known, data are available indicating which themes have innovation projects within and outside the MTIB and a counterfactual can thus also be designated. The latter must nevertheless be very clear as to what the MTIB's prioritised solution pathways are, and how innovation projects within the associated schemes (NWO KIA calls, PPP allowance, MIT, MOOI) and outside those schemes (WBSO) relate to those directions. For the ultimate comparison, projects must be very precisely labelled, as this shows whether chosen innovation directions in the field of sustainable manufacturing, for example, are now developing faster and better than the directions that were not chosen. A promising point in this regard is the fundamental principle that the MTIB should provide clear directions for combining innovation and diffusion efforts. The missions themselves are only the ambitious goals, but the governance structure and policy efforts and instruments are aimed at giving momentum to the most promising solutions. Those directions are explicitly documented in the KIAs and MMIPs and serve as a basis for the programming of various schemes.

If the evaluation does not make clear which directions are involved, that is at the outset a very bad sign for the effectiveness of the policy. Moreover, the implication with regard to the selection of empirical methods would be that alternatives with a lower causal evidential value should be adopted.

The flowchart refers in this case to contribution analysis or process tracing. Its application would amount to a structured review of indications that there is a series of desirable developments, consisting in any case of: more thematic and organisational/administrative coherence between potentially complementary policy instruments and efforts; thematic convergence of the innovation projects supported by various schemes (in relation to the MMIPs) and an accelerated and more successful development of the projects related to the MMIPs (in terms of patents, publications, prototypes), ideally also in relation to other innovation projects.

The Climate Policy

Description of the policy

The objective of the Climate Policy is to reduce emissions of greenhouse gases. The recent coalition agreement includes much more ambitious objectives, and is intended to achieve climate neutrality by 2050. An interim goal of a 55 percent reduction by 2030 has been defined. These goals are laid down in the Climate Act. The Climate Plan 2021-2030 has been drawn up on the basis of this Act. This Climate Plan is largely determined by the key points of the Climate Agreement. The Climate Plan provides a comprehensive picture of the Dutch government's climate policy.

The ultimate objective requires not a single transition but multiple transitions simultaneously in the various sectors and production-consumption systems. Moreover, these transitions do not all take place in the same way, and they have different 'buttons' that need to be turned. What works in one transition will not work in the other.

There are no quick fixes to achieve these transitions. Fossil energy is deeply interwoven with the sociotechnical and production-consumption systems on which Dutch society relies. Climate neutrality requires transformations of these systems, for which technical, societal and organisational innovation is required. The innovation policy is aimed at achieving these socially desirable innovations and thereby 'feeding' transformations (long-term processes).

The Climate Policy is assigned to a number of departments, under the ultimate responsibility of the Ministry of Economic Affairs and Climate Policy. The Climate Policy concerns a multitude of schemes, instruments, laws and activities. Although the innovation toolbox of the Ministry of Economic Affairs and Climate Policy is central in this illustration, there are (sometimes strong) interfaces with other instruments. Some examples of instruments are:

- Energy innovation schemes, TSE, MOOI, DEI+, HER and VEKI;
- IKIA Climate and Energy with associated MMIPs;
- Energy Investment Allowance;
- SDE++ as a subsidy for renewable electricity projects;
- Structure and implementation of the Regional Energy Strategies (RES);
- 'Greenhouses as Energy Source' programme;
- Implementation of the National Charging Infrastructure Agenda;
- Green Deal on sustainability of waterborne transport;
- Amendment to the Environmental Protection Act (businesses must take energy conservation measures that pay for themselves within five years);
- Natural Gas-free Districts Programme (PAW);
- Prohibition of coal-fired electricity generation from 2030.

European policy instruments also play a role in Dutch climate policy, such as the ETS (where the Netherlands conducts additional policies, for example through the Subsidy Scheme for ETS Indirect Emission Costs and the CO_2 levy). In addition, the climate policy touches on a number of other overarching policy areas, either because the policy areas to some extent implement each other's policies or because there is tension between the policy areas.

There are a number of reasons for the government to intervene in the climate transition. These reveal the main issues that need to be resolved. The main reasons from the perspective of system and transition policy are:

- *Direction* – Translating the climate objective within sectors is essential to avoid initiating incompatible change processes or to avoid a failure to invest in those changes because it is not clear which new system the operators should anticipate (directionality failure). Setting clear goals and defining solution directions should resolve this. Formulated subgoals must in turn be embodied in actions assigned to specific parties that make measurable contributions to the specified objective;
- *Speed* – Fundamental changes in production-consumption systems proceed too slowly without government intervention. The importance of speed becomes clearer if we consider that tipping points can be reached in temperature rise, with consequences becoming irreversible;
- *Policy coordination* – Changes in production-consumption systems affect so many interrelated areas (e.g. in intertwined value chains in which synchronised adjustments are required in production, transport and use of goods and services), that policy coordination is required.[29] This requires close and continuous alignment between policy and practice, and among the policy areas, in order to address the practical issues that arise;
- *Missing system functions* – Innovation systems intended for climate-neutral production-consumption systems do not always function properly. For example, due to restrictive 'fossil' regulations or a lack of entrepreneurship. Market failures also play a part in the changing system, e.g. because negative effects are not priced in or even because non-sustainable commodities (such as fossil energy) are subsidised.

In addition to implementing all kinds of instruments to promote innovation (and its adoption), the Climate Policy includes control mechanisms to introduce coordination and harmonisation into the transition.

<u>Considerations based on the decision aid</u>

For the Climate Policy we go through the steps in the decision aid in Chapter 7. We provide a number of considerations in each of the steps or questions. The starting point is that we are dealing with policy featuring a multitude of instruments and activities pursuing a single policy goal.

Are the instruments synergistic?

One of the objectives of the overarching climate policy is the introduction of coordination of policy, instruments and activities. New and existing instruments that simulate innovation are brought together under the heading of climate policy. A number of these instruments were already synergistic in terms of design (e.g. HER and SDE+). Other instruments have also been added, including from other departments. This creates a range of instruments and activities that aim to support innovations from idea to implementation, with broad attention being devoted to co-innovation (e.g. organisational or societal innovation that is also necessary).

In terms of their intent, the instruments appear to be focused on the same objective. For most instruments, however, it is still unclear whether and how synergy has been achieved (including in policy design). It is therefore advisable to see precisely how this works out. For the climate issue, the obvious way forward is therefore to investigate first how processes work to achieve synergy (governance) and whether instruments are also synergistic in terms of design (policy mix analysis); see next page.

[29] This is related to the chosen varied mix of instruments and polity. The fewer the instruments used, and the more central control or influence there is, the less policy coordination is required. This is related to potential *government failure*, which occurs if the government lacks the capacity and knowledge to coordinate (possibly complex) policy properly.

The above does not alter the fact that a number of the instruments will also have to be evaluated individually for effectiveness and efficiency. The Ministry of Economic Affairs and Climate Policy has taken steps to carry out individual instrument evaluations to increase comparability. In this way, individual evaluations can contribute to an evaluation at system level.

Once it is known whether governance contributes to synergy in instruments and activities, and instruments are actually synergistic in design, the decision aid can be gone through again. In this example we therefore consider a number of follow-up questions.

Is there a measurable policy objective?

The Climate Policy has measurable and quantifiable policy goals for 2030 and 2050. The innovation toolbox also play a part in this, albeit mainly in the phases of the transition in which innovation is an important driver for completion of the phase. It is recommended that these policy objectives be closely monitored.

There are also intermediate goals for the various production and consumption systems, for example that all new passenger cars must be 100 percent electric by 2030, that the electricity network must become flexible or that industry must be able to act as a buffer for energy networks. For such production and consumption systems, it is possible to examine whether the innovation system functions optimally and whether changes actually occur in the structure of production and consumption systems.

To determine whether those changes are caused by policy, it is always necessary to identify which of the many possible policy instruments have influenced the system in question. It is possible that some instruments act on multiple systems. This does not mean it is sufficient to evaluate those individual instruments; after all, the question is whether the interaction of all those instruments (and additional coordination) actually strengthens or weakens a system. The ideal evaluation design and the methods to be used will probably differ from system to system. We will not go into further detail in this summary analysis.

Are the system issues known? Can the system be clearly delineated? → *System analysis*

If there are no clear intermediate goals, an evaluation can focus on examining the relationship between policy and remedying issues impeding the transition steps. The question is primarily whether those issues are already known. As part of the monitoring of the Climate Policy, the Ministry of Economic Affairs and Climate Policy is working on studies in which various 'sectors' (and their production-consumption systems) are examined to see how they stand. A workable delineation of sectors has been found that makes it possible to examine whether improvements will also take place over time. The decision aid suggests doing this by means of system analyses, for example on the basis of the analytical framework in which the various key processes of technological innovation systems are compared.

In terms of methods, it is appropriate to opt for contribution analysis or (where concrete hypotheses can also be formed on the basis of policy theory) process tracing based on empirical material that can provide insight into all activities conducted in the Netherlands in the selected sectors / systems. This can include an examination of the data that the Netherlands Enterprise Agency maintains when implementing many of the aforementioned schemes. Of specific interest is the coherence across these schemes, since some schemes provide more insight into knowledge development, while others focus on cooperation, marketing sustainable innovation, integrating partial solutions, etc. By bringing together the records of various schemes, it is possible to provide a comprehensive picture of which actors/networks collectively make up the system, and what emphasis is placed on strengthening those systems. Subsequently, the essence of contribution analysis and process tracing, including on the basis of additional material, is to examine step by step which data indicate the sequence of changes that should occur. For example, if the system surrounding the modification of the electricity network encounters an issue of lack of knowledge, we must ensure that additional resources are devoted to

the specific subject on which knowledge is lacking, that additional projects have been established and that this in turn has led to concrete knowledge results (articles, reports, patents etc.) on that subject. For an issue relating to behaviour, for example, the causal chain will probably be more complex, which emphasises how important it is first to reconstruct the steps through which interventions and results should ultimately deliver impact.

→ *'Study governance and policy mix'*

When policy theory is not clearly defined, and hence it is unclear how policy should help eliminate issues, it may be beneficial to first gain a better understanding of the actual organisation of the governance and the design of the policy mix – including which factors were key and whether they were also implemented. There is also a component in the monitoring approach to Climate Policy that requires insight into whether the examination of and response to issues is properly organised. Understanding this requires an analysis of the governance of the climate transition and of the mix of policy instruments that aim to boost it. This raises the following questions: Are transition processes running adequately (governance, cooperation, follow-through, policy coordination etc.)? Are the processes in line with principles? Are the plans and instruments coherent and consistent?

In the evaluation of **governance**, the question is whether control frameworks and principles have been formulated and introduced that ensure that the climate transition (across all kinds of connections in society) is embodied in specific transition processes. The value of examining this hypothesis is twofold. First, it is currently unclear whether the arrangements in place with regard to management principles and management activities are actually functioning as expected; managing transitions is still relatively unknown territory for governments and researchers. In addition, such an analysis provides insight into the way in which objectives such as the acceleration and direction of the transition are achieved. Both types of insight can be of value in improving the overarching transition policy.

Policy mixes have often grown historically. In the case of climate policy too, instruments are being brought more into line with the objectives of the policy. In addition, 'old' policy continues to exist. New instruments and actions are also being developed to solve additional issues. These policy initiatives are taken at both regional and national level. Overall, this results in a multiplicity of policy incentives in an extensive area such as Climate Policy, with analyses not necessarily being carried out in advance on the policy mix as a whole. This leads to the question of whether the policy mix relating to the climate issue is optimal. This can be examined by looking at the actual scope/delineation of policy documents and regulations (e.g. the tenders), the views of stakeholders on these, and the portfolio of consortia/projects that move 'due to the policy mix'.

Examining governance and the policy mix is not so much a question of finding the right empirical methods (it often comes down to interviews and desk research), but above all of asking the right questions. This involves finding 'assessment frameworks' that can serve as a mirror for the actual governance and policy mix. Such assessment frameworks can be derived from both the (scientific) literature and discussions with those involved who shape and implement the policy. We therefore propose a step-by-step plan for both perspectives for:

1. Identification of the initial design and principles of the governance and policy mix;
2. Reviewing the scientific principles around governance and an optimum policy mix;
3. Description of the practice, both of control processes and of the policy mix.
4. An analysis of practice versus the theory.

Reflexive evaluation can be of use in these analyses as a research method. In doing so, evaluators and practitioners give meaning to both the evaluation 'yardstick' and the performance against that yardstick.

Box 1. Reflexive evaluation as a means of evaluating the governance process.

We suggest approaching the evaluation of transition governance according to the principles of 'reflexive evaluation'. The background to this approach is described *inter alia* in the Policy Evaluations Toolbox[30]. This approach fits in well for the following reasons:

- It is not clear what the ideal method of governance is for a transition issue such as climate. Scientific principles provide guidance, but the evaluation lens needs to be further refined and operationalised. 'Good' governance is determined jointly with practitioners.
- For the Climate issue itself, it is not yet clear exactly how the transition governance will work out in practice and how it should be further developed. A reflexive evaluation will provide support for this process of professionalisation;
- The transition is a process that remains in development. This dynamic lends itself best to an *ex durante* approach to evaluations, involving regularly assessments of how the process is developing and could be adjusted.

Five (general) phases can be distinguished in the evaluation process for reflexive evaluation (partly inspired by the work of PBL):

1. Formulating evaluation questions or sub-hypotheses in consultation. This can be done with the parties involved, such as the departments and the implementation committees for the Climate Agreement;
2. The joint development of an evaluation framework – this corresponds to steps 1 and 2 presented above, in which we consider jointly with the stakeholders which frameworks and principles are fundamental for the management of innovation in the Climate Transition. The further development of the policy theory fits in with this;
3. Data collection (mixed methods) – this corresponds to step 3, and mainly consists of desk study and (group) interviews. Quantitative data, such as data gained during monitoring, can also serve as input;
4. Joint interpretation – in a workshop setting, stakeholders can work with the researchers to give meaning to the factual material found, giving an interpretation of the progress, operation and vulnerabilities with regard to governance;
5. Formulating courses of action – this is what the entire learning process is mainly aimed at, namely embedding lessons in the policy context, already during the research. This means developing courses of action that policymakers will actually get to work with.

Reflexive evaluation is relatively new. Stakeholders are therefore often not yet alert to the possible pitfalls or issues associated with this method. Possible issues and mitigation measures in a reflexive evaluation are:

1. The time required from stakeholders is substantial – reflexive evaluation is relevant to implementation practice, and hence also to the stakeholders' practice;
2. Stakeholders need to demonstrate openness – a professional and personal attitude on the part of the researcher is important. A setting of trust should be created;
3. The parties lobby intensively, so good researchers are needed who can supervise the research process and allow all voices to be heard equally, while also challenging lobbyists with critical questions. This may also mean the researchers are able to draw the conclusions from their evaluation independently of the stakeholders;
4. Doubt is cast on the independence of researchers (and the validity of conclusions), so clear assignment and separation of roles is important. The process must be designed transparently, and it must be clear how conclusions and courses of action have been arrived at.

The Dutch Digitisation Strategy

Description

The Dutch Digitisation Strategy (NDS) is a government strategy (led by the Ministry of Economic Affairs and Climate Policy) that combines ambitions and objectives in various policy areas for a successful digital transition in the Netherlands. The policy consists of a collection of all the policy that is conducted by these ministries, and that can be grouped thematically under the agenda.

The main objective of the NDS is to have a single strategy for the Netherlands. The first-order effect of a single strategy, rather than a disjointed collection of policy activities, is a strengthening of the exchange of knowledge and best practices, avoiding reinventing the wheel or introducing conflicting regulations in different places.[31] The second-order goal in terms of outcome as formulated in the letter is threefold: we become Europe's digital leader; everyone participates and benefits; and there is confidence that digital technology is being used with care. The NDS takes concrete form as follows:

[30] Toolbox policy evaluations – research methods – 011. Reflexive evaluation
https://www.toolboxbeleidsevaluaties.nl/onderzoeksmethoden

[31] https://www.nederlanddigitaal.nl/english/dutch-digitalisation-strategy

taking advantage of social and economic opportunities and strengthening ground-breaking research innovation, new skills and lifelong learning, a dynamic digital economy, digital resilience of citizens and organisations, and establishing fundamental rights and ethics in the digital age. The ultimate goal of the NDS in terms of impact is greater prosperity through economic growth and safeguarding public interests.

The strategy is updated annually, with the subgoals sometimes shifting or becoming more precise. The policy includes actions by and activities of the following ministries: Economic Affairs and Climate Policy; Interior and Kingdom Relations; Infrastructure and Water Management; Health, Welfare and Sport; Agriculture, Nature and Food Quality; Education, Culture and Science' Social Affairs and Employment; Defence and Justice and Security. Various action plans are also included in the policy: Examples include the Dutch Cyber Security Agenda (DCSA); the Digital Connectivity Action Plan; NL DIGIbeter, Digital Government Agenda; NL DIGITAAL, Government Data Agenda, SME Action Plan and Digital Inclusion Action Plan, Strategic Action Plan for Artificial Intelligence. In addition, the digitisation strategy includes a number of coalitions: Dutch Blockchain Coalition, collaboration platform for cybersecurity knowledge and innovation, Dutch AI Coalition, Online Trust Coalition Data Sharing Coalition, digital society alliance.

The overview of financial resources for the digital economy lists almost all generic instruments of the Ministry of Economic Affairs and Climate Policy (the PPP allowance, international innovation, innovation credit, EuroSTART, MIT, and the WBS, SBIR), plus a number of specific instruments (Smart Industry Fieldlabs and SME action plan, Thematic AI calls and projects, AI expenditure by ministries other than Economic Affairs and Climate Policy and Education, Culture and Science, AI investments by European funds, and AI investments by the National Growth Fund).

The most recent report on the digitisation strategy seeks to provide an overview of results (and thus implicitly the objectives pursued) by attempting to substantiate a number of propositions: (1) The Netherlands is high in international rankings in the field of innovation and digitisation; (2) digitisation helps meet societal challenges; (3) digitisation contributes to economic growth; (4) there is increasing cooperation in the field of digitisation; and (5) more attention is paid to the governance of digitisation.

Regardless of whether the available data supports these statements, the question is whether this is the intended effect of the policy ('the digitisation strategy'). The link between the overview and the intended objectives is unclear. Policy efforts can also be directed to the policy on the basis of this overview, and the question is what exactly the digitisation strategy is achieving in these areas. A relevant question in such a summary is of course the extent to which the performance has been achieved thanks to the strategy: performance may also have been achieved without policy or may have been driven by market factors.

The digitisation strategy is *de facto* bringing together under one heading various policy activities that have already taken place and been deployed. It is not clear how these activities have been adjusted or strengthened as a result of the digitisation agenda. A complicating factor is that the intended objectives of the NDS are formulated in terms of the original objectives of the components that make up the agenda. As a result, the objective of the strategy itself as a policy instrument disappears from view.

<u>Considerations based on the decision aid</u>

The digitisation strategy consists mainly of a number of combined policy initiatives and a coordination mechanism. A number of these initiatives can be individually assessed for effectiveness and efficiency, because they are instruments with a clear objective and budget. At first sight, this aspect of the NDS is clearly formulated and verifiable. However, data must be provided to verify the objectives. There is often no control group available, but it would be possible to examine whether policy has been

deployed regionally or for different businesses/sectors at different times. If that is not the case, an effect can be demonstrated using, for example, more qualitative methods or regression analysis.

However, there are also components that are not clearly formulated and verifiable. An effort must be made to clearly formulate the instruments within the NDS on the basis of objectives and resources in order to evaluate them. Regression techniques can then be used to understand the extent to which there are indications of the effectiveness of the instruments. Probably only an explanatory analysis is possible at the moment.

Finally, the coordination of the activities in a single strategy should also be evaluated. It is possible to evaluate whether there has been greater collaboration since the coordination and whether synergy between instruments has been improved. This can be verified by means of regression analysis. It is also possible to assess the efficiency by analysing the extent to which the same objectives are achieved with fewer resources after the establishment of the NDS.

The Technology Pact

Description

The aim of the Technology Pact is to increase the number of technicians and to strengthen the scientific, technical and technology skills in the Dutch labour market. It is designed on the basis of choosing, learning and working. Choosing concerns early interest in and embedding of technology in education, from basic through to vocational. Learning concerns collaboration between businesses and education to keep teachers' knowledge up to date, create sufficient internships and apprenticeships and foster greater cooperation between higher education and business. The work focuses on permanent learning, the retention of technology workers and better collaboration between national and regional levels.

The Science and Technology Platform foundation (*Platform Bèta en Techniek*, formed on 9 July 1998) aims to reduce the science-technology problem by increasing the attractiveness of science-technology in education and the profession, thereby promoting cooperation between intermediaries and actors in education and business. The Technology Pact is a part of this and consists of a partnership of educational institutions, employers and employee organisations, regions and central government. The Technology Pact operates as a network which different parties can join and in which initiatives are developed and implemented at national and regional level. The objective is to resolve issues in choosing, learning and working. Detailed agreements were entered into in 2013 by businesses, education and government for a period of seven years. Those agreements were redefined and updated in 2018. The principal aim of the renewed pact was to link the Technology Pact to new societal challenges, such as the energy transition, climate change and the circular economy. A new strategy was drawn up in 2021. The ambition of the Technology Pact is to adopt a structural approach in order to contribute to a well-trained workforce with sufficient technicians for the jobs of today and tomorrow. The context of this ambition is periodically adjusted, because the purpose is clear but the demand for scientific, technical and technological skills is constantly changing.

The Technology Pact is being implemented in five regions: North-west, North, East, South-east and South-west Three ministries are involved: Education, Culture and Science, Social Affairs and Employment, and Economic Affairs and Climate Policy. The Technology Pact operates as a network which parties can join. The objectives and returns are not specified by central government, but depend primarily on tackling issues affecting demand and supply in the region.

The National Management Group consists of a number of parties, such as representatives of the government (ministries of Economic Affairs and Climate Policy, Education, Culture and Science, Social Affairs and Employment), of the five regions, of the Talent for Technology platform, of education & training councils/associations, sector/employer and employee organisations, the Top Sectors, the UWV benefits agency and partnerships/platforms such as the Foundation for Cooperation on Vocational Education, Training and the Labour Market (SBB) and the Platform for the Information

Society (ECP). The National Management Group is led by the chairman, who also acts as a figurehead for the Technology Pact.

The national level is then divided into five regional sections: North-west (Noord-Holland, Utrecht, Flevoland), North (Friesland, Groningen, Drenthe), East (Overijssel, Gelderland), South-east (Noord-Brabant (part), Limburg) and South-west (Zuid-Holland, Zeeland, West-Brabant). Each region has a coordinator who participates in the National Management Group. The coordinator is often at the level of a member of the Provincial Executive, alderman or portfolio holder. For each region there is also a central contact person on a more operational level. These take part in joint consultations within the liaison committee.

The Talent for Technology Platform supports coordination between the regions and activities in the regions. The Talent for Technology Platform is a public-private partnership in which the ministries of Economic Affairs and Climate Policy, Education, Culture and Science and Social Affairs and Employment collaborate with employer and employee organisations and eleven sector/employer and employee organisations such as FME, BOVAG, De Unie and FNV. Among other things, the Talent for Technology Platform builds on the old Science Technology Platform, TechniekTalent.nu and TecWijzer. In addition to the resources for the annual conference, the financing of the work of the Talent for Technology Platform is one of the few direct financial investments by central government in the Technology Pact. The Talent for Technology Platform has one central contact person for each region.

The action lines consist of objectives with specific implementation in each region. The National Technology Pact Control Group (LRT) coordinates, follows and monitors the implementation of the national strategy, the objectives and the agreements made in the Technology Pact. The management group comprises representatives from the five regions of the country, central government, employers, employees, top sectors and education. The ministries of Economic Affairs and Climate Policy, Education, Culture and Science and Social Affairs and Employment are responsible for the national actions within their fields. The Technology Pact project team is based in the Ministry of Economic Affairs and Climate Policy.

A number of organisations are involved.

- Strong Technology Education. The Ministry of Education, Culture and Science, Stichting Platforms vmbo (SPV) and the Talent for Technology Platform (PTvT) will work together to promote Strong Technology Education in the years ahead. They will work closely with VNO-NCW, the Federation of Technology, MKB-Nederland, Technical industries, Platform TL, Dutch VET, and VO-raad. The national Strong Technology Education support team helps schools and regions to make and implement regional plans. They provide support, data, inspiration, tools and information to help regions establish the partnership, the planning process and implementation.
- Regional secondary/higher education networks. Nine networks of secondary and higher education institutions with the aim of continuous subject and curriculum development in secondary education, improving the links between secondary and higher education and continuous professional development of teachers, teaching assistants and school leaders in secondary and higher education.
- Katapult is a network of more than 300 partnerships between education and business and is constantly growing. The objective is to improve cooperation between education, business and society. This is done, for example, by professionals from the world of business who provide lessons. Or by students conducting research for an SME during their educational programme. There are now 84,000 students, 9,800 businesses and 5,000 teachers in these partnerships, also known as Centres of Expertise, Centres for Innovative Professional Skills and other forms of collaboration in vocational education. There is an annual subsidy of around €0.3 million for this organisation.

- Jet-Net & TechNet. This organisation fulfils the joint ambitions of the technical employer and employee organisations, the government and the business community to jointly organise guest lectures, company visits and projects. There is an offer for primary education, secondary education and businesses. The partners are Bouwend Nederland, BOVAG, FME, Koninklijke Metaalunie, Techniek Nederland, CNV Vakmensen, De Unie, FNV Metaal, VHP2 and the Ministry of Education, Culture and Science. There is an annual subsidy of around €0.8 million for this organisation.

Four themes were identified in 2021. The total subsidy for the Technology Pact is around €0.8 million per year. The total subsidy for the Science and Technology Platform foundation is around €3.1 million. The themes below contain the government's approach to the coming period, for which the ministries are responsible.

- Inflow
  - girls/women more often choose a technical programme and job
  - young people with a migration background more often choose a technical programme and job
  - young people more often choose a technical intermediate vocational education programme
  - embedding knowledge and digitisation in the primary and secondary education programme
  - increase in the number of apprenticeship businesses
  - close cooperation between education and business in the field of higher vocational education

  The implementation, support and coordination of initiatives, activities and programmes takes place at regional level, including through Jet-Net & TechNet (primary and secondary education), Strong Technology Education (pre-vocational); regional secondary and higher education networks (senior general or pre-university) and Katapult (intermediate vocational/higher vocational).

- Public-private partnership
  - strengthening practical education and cooperation in technical higher vocational education
  - scaling up good initiatives
  - drawing attention to financing schemes and possible applications

  Katapult brings together all Dutch public-private partnerships.

- Shortage of teachers and professionalisation
  - give hybrid teaching a place on regional agendas and in sector plans
  - teacher training courses should be more closely aligned with the knowledge and experience of career switchers

  The implementation, support and coordination of initiatives, activities and programmes takes place at regional level, including through Strong Technology Education (pre-vocational); regional secondary and higher education networks and Katapult (intermediate vocational/higher vocational).
- Lifelong development
  - increase in numbers retraining for and upskilling in the technical professions
  - drawing attention to financing schemes and possible applications
  - safeguarding prospects in technical professions in digital provincial lifelong development programme portals

Monitoring

- The Technology Pact monitor provides relevant data on technical training courses, labour market shortages etc. and is constantly updated. The information is also broken down regionally, making the monitor user-friendly for policymakers in the region. See www.techniekpactmonitor.nl.
- The Talent for Technology Platform (PTvT) is tasked annually by the three ministries to organise the links with and between regions in the Technology Pact. The subsidy assignment also includes communication and events. PTvT reports annually to the ministries on the progress of the activities.
- The three ministries report annually to the House of Representatives on the activities and progress towards the objectives within the Technology Pact. Up until 2019, this was done by means of a progress report detailing a wide range of activities (including regional). At the end of 2020, it was decided to send an agenda with future actions to the House in advance for the 2021 calendar year. This is in line with the Technology Pact evaluation by SEO Economic Research and Technopolis, which recommended a more programmatic approach to the work. An initial accountability process for this government commitment in the new form has not yet taken place and is planned for late 2021 / early 2022.
- As mentioned, the Technology Pact can be seen as a network approach, a kind of umbrella under which different organisations arrange different activities in line with the Technology Pact objectives. For example, schemes such as MKB!dee (€40 million) and the RIF (€100 million) should be regarded as instruments that fall under the umbrella of the Technology Pact policy. These schemes are evaluated independently, however.

Considerations based on the decision aid

The Technology Pact addresses a number of issues in education and in the labour market for technically trained personnel. The vision (from 2013) is that a structural and overarching approach contributes *to a well-trained working population with sufficient intelligent and competent technicians for the jobs of today and tomorrow.* The central failure is a failure of capacity in education. From primary education through to the labour market, insufficient attention is devoted to knowledge of technology. As a result, children do not come sufficiently into contact with technology, too few children opt for technical education and the children who do often switch. As a result, bottlenecks arise in the labour market and business innovation stalls. Capacity failures also exist in the labour market because there are too few skilled teachers who educate children and insufficient investment is made in new knowledge and skills to retain workers in technical professions or improve their employability.

The low inflow into technical education is the result of misconceptions about technical education and professions on the demand side and insufficient supply of effective education. Problems on the supply side can be subdivided into a lack of sufficiently qualified teachers in primary and secondary education and subsequently in technical education and insufficient involvement of the business community, resulting in a sub-optimal match between education and the labour market.

There is insufficient training in technology once people are in work. On the demand side, the issues are as follows: (i) restrictions in the education or financial markets may provide an explanation for inadequate training, (ii) external effects because workers who would benefit most from training do not (or cannot) take advantage of the education on offer or have no incentive, (iii) because of the holdup problem (the employer has no guarantee that the employee undergoing training will continue to work for the company in the future and the employee who continues to develop will be uncertain whether the benefits will lead to higher pay), and (iv) explanations relating to behavioural economics, such as short-sightedness, loss aversion, defaults and norms, that lead to a sub-optimal learning

culture. On the supply side, there are limitations in the range of courses in the education market due to, for example, high investment costs for starting courses and market distortions due to cross-subsidisation of educational institutions.

An action agenda has been drawn up this year (2021), enabling the government to chart a direction. In practice it involves much more than that. At provincial, labour market and municipal level, education, business and governments work together on technical education and jobs in technology, construction and ICT. This regional cooperation is the foundation of the Technology Pact approach and will continue. At national level too, there are many initiatives and instruments that are important to achieve the objectives of the Technology Pact. The actions on the agenda for 2021 can be evaluated.

The inflow into technology can be cited as an example. A number of concrete objectives and instruments have been defined. Together these should ensure that the inflow is increased (output), leading to more economic growth and greater competitiveness (outcome).

The instruments and activities associated with this are focused more on girls and women in technology:

- Development of the Science & TechMentality model. In order to inspire girls about the potential of technology, it is important first to know what drives and motivates them;
- Diversity check. Dealing consciously with differences between boys and girls;
- Communication check. Gearing communication better to the female target group;
- 30 percent of women in technology in 2030. FME has the ambition to more than double the number of women in the technology industry to 30 percent by 2030 by making it more attractive to work longer hours and tackling gender stereotypes.

More young people with a migration background in technology:

- Increase the visibility of technology;
- Ensure widespread familiarity with technology;
- Increase technical self-confidence;
- Emphasise the potential of a future with technology;
- Develop a clear policy on discrimination in technology.

Focus primary and secondary education more on technology:

- Structural commitment and approach across all educational levels, phases and chains. Technology-related education will be included in the revision of the curriculum;
- In the short term, institutions can use Strong Technology Education and start working with science & digitisation from Jet-Net & TechNet.

Higher inflow into training-on-the-job pathway: More students must opt for technical courses, because the labour market has a growing need for people with a technical intermediate vocational diploma (including training-on-the-job).

Strengthening practical education in higher vocational education: The practical component in higher vocational education should receive more attention. The 'SME route in higher vocational education' programme has started with seven pilots

Attracting international talent: The Talent Coalition is working on positioning the Netherlands as an attractive location for international talent in paid employment.

It is a substantial challenge to evaluate the Technology Pact as a whole (see SEO/Technopolis, 2020). The network approach and the resulting output can nevertheless be monitored and assessed for

consistency. Specific subgoals (such as the inflow into technical training throughout the chain) can be evaluated. Finally, it is possible to carry out evaluations at instrument level (e.g. higher inflow of girls into technology or programmes such as the Regional Investment Fund).

At system level it is unclear what the optimum number of people in technology is. As a result, it is unclear how big the impulse should be in education with regard to adjustments in the curriculum (if technology receives more attention, that will be at the expense of another subject) and with regard to influencing young people's choices. It is nevertheless possible to measure the extent to which issues arise and instruments can be deployed and evaluated.

Young people make choices based on preferences, their network and previous experiences in education. In that way they discover their optimum educational programme. Adjusting the educational paths is only useful if it is established that young people have made suboptimal educational choices. If young people do not want to work in technology, stimulation is not effective.

Adjustments to the curriculum will probably take longer than stimulus measures to encourage people to choose technology. For choices to be effective, it is important that both are not too far apart in time.

It is possible to measure how many boys and girls with a migration background choose technology. The same applies to the inflow into the training-on-the-job pathway, practical education in higher vocational education and attracting talent from abroad. However, it is necessary to define the precise purpose and the period within which this should be achieved. A change in the curriculum should be determined and its effects on the inflow should be measured.

The various initiatives can be implemented in concrete form. Efforts to increase the number of girls in technology appear the most concrete. By running specific programmes in a number of regions, it is possible to measure the extent to which these efforts are effective and efficient and it is possible to learn to what extent programmes could also work in other regions.

The coherence along the education chain can be measured over time. If programmes are introduced in primary education to influence choices, it is possible to measure in secondary education the extent to which more girls and boys with a migration background opt for a technical profile and then choose a vocational education programme in technology. It is also possible to conduct interventions here and assess their strengthening effect.

It is certainly possible to measure the effects of the programmes and instruments that fall under the umbrella of the Technology Pact in terms of effectiveness and efficiency. A precondition is that concrete objectives are set and programmes and instruments are not introduced simultaneously in every region so that effect measurement is possible (causality can be determined). The data are available, particularly with regard to the inflow and progress through education. The societal impact of the Technology Pact as a whole is not hard to determine. The measurement is limited to consistency of approach.

8.2. Lessons on the delineation of (and designation as) S/T policy

If we look at the similarities and differences between the four cases examined, the first observation is that delineating the policy fields to be evaluated is a delicate matter. For example, a policy area such as CO_2 reduction policy or climate policy does not correspond to a concrete policy issue (with instruments and associated budgets) and the Mission-oriented Top Sector and Innovation Policy does not correspond to the way in which some documents and stakeholders view mission policy. The National Digitisation Strategy is very broad and consists of a combination of usually existing instruments implemented by various ministries. Finally, the Technology Pact has existed somewhat

longer and the emphases have changed over the years. Since S/T policy is not about an instrument but about a collection of policy impulses that fall within a possibly formalised programme, it can be difficult to get to grips with the policy strategy and the policy package that needs to be evaluated; see also step 3 in the step-by-step plan. Alternative delineations also affect other objectives, which means questions soon arise as to what evaluations should actually focus on. The client commissioning an evaluation should eliminate such uncertainties at an early stage, since ambiguous principles can impede the development of evaluation designs to answer the most pressing evaluation questions and lead to a debate on what should now actually be evaluated. This immediately raises the question of how to prevent the client from using demand-driven processes to determine to an excessive extent what can and cannot be evaluated. After all, by making informed choices about what and how to evaluate, a client can prevent overly critical assessments. This is a type of government failure that affects the design of evaluation processes within the government itself, which is outside the scope of this research.

A second observation concerns the variety of the cases examined, specifically with regard to the characteristics that are distinctive for S/T policies. The MTIB and the CO_2 reduction policy probably come closest to the interpretation of S/T policies used in this report, as they are overarching strategies that themselves rely on interventions (including in innovation) that should act on the same system in complementary ways. But there are also cases that do not have these characteristics. We distinguish three cases:

- One category consists of policies where the overarching strategy is less central, and where the emphasis is on a collection of concrete policy interventions that do not claim to be very synergistically related. A similar factor plays a role to a certain extent in the National Digitisation Agenda: it is more like a thematic overview of policy impulses than an integrated strategy with synergy between the various components. A consequence of this policy design is that in such a case it is better to evaluate each individual instrument on the basis of the standards of the Effect Measurement Expert Working Group (2012), as there is little coherence;

- There are also ostensibly S/T policies that place much greater emphasis on the overarching strategy or programme; although there is a collection of concrete policy interventions, the relationship with a concrete set of policy interventions is relatively weak. An example is the Smart Industry programme or (the current state of) the policy strategy on the Circular Manufacturing Industry. In both cases, the policy emphasis (in terms of resources) seems to be more on scheduling and coordination than on deploying or significantly adjusting concrete forms of government intervention. The implication for policy evaluation in this case is that it is probably best to focus only on perspectives A or B from the evaluation framework, since the other perspectives also require there to be substantial incentives through which system strengthening/transformation (and subsequent outcomes) can be achieved;

- Finally, it is also possible that in the case of S/T policy there will be no overarching coherent strategy, and actually also no package of heterogeneous policy interventions. This appears most relevant in the case of the Technology Pact, which primarily relies on organising a network in which various parties work together to increase scientific, technical and technological skills in the Dutch labour market. Although systems thinking is used here and developments and progress are charted annually, the strong focus on instruments or delineated subgoals shows that evaluations of each instrument are probably also possible.

The general lesson is that not every policy agenda with system or transition aspects automatically corresponds to the S/T policy type on which this report focuses. It is possible that the evaluation options considered here are relevant to certain components, for example with regard to the

evaluation of governance processes and structures, but it is still sensible to look critically at whether an approach such as that proposed by the Effect Measurement Expert Working Group (2012) is also among the possibilities. This is particularly the case if the S/T policies include individual instruments acting on clearly delineated subgroups of actors.

8.3. Lessons across research methods

There are also several lessons to be learned on the basis of the research methods to which the decision aid refers after going through the flowchart. In the cases in which quantitative effect measurement appears possible, it is a question of studying structural changes brought about by the Technology Pact and NDS, because those cases comprise relatively independent interventions to which the methods in the report of the Effect Measurement Expert Working group (2012) lend themselves (see Section 6.1). However, methods with high scores on the effect ladder can also be used in the case of the MTIB, which is more typical of S/T policies. The following observations are important here:

- Experimental or quasi-experimental methods are applicable to sub-aspects of the S/T policies. Chapter 3 already considered why, for an S/T programme, it is often difficult to designate a counterfactual, for which relevant data must also be available. As a result, it is difficult to determine whether MTIB policy, e.g. for the circular economy, is responsible for improved performances on the latter point, or whether positive performances are attributable to other developments. What appears more feasible is the search for policy effects on specific intermediate goals, such as structural changes comprising changed investment or production plans (which should be the result of system strengthening). An evaluation of MTIB policy could use the fact that the MTIB seeks to strengthen research aimed at specific solutions, and that, for example, developments in innovation efforts on the selected themes – as observed in WBSO data – can be compared with developments in similar sectors or technology fields that are *not* prioritised on the basis of the MTIB policy. Such an analysis does not show whether that extra innovation effort also results in more circular economy, but it does give an indication of possible crowding-in as a result of system strengthening that is inherently difficult to measure.

- In the case of S/T policies, there is a need for data at a very specific level of detail concerning the direction of all kinds of investments and efforts for which there are often no ready-made answers. In the case of simple instruments aimed at individual actors, a treated group and a control group can be distinguished, but S/T policies revolve around developments at the meso level that cannot easily be traced back to individual actors. This means measurements must be made at the level of abstract concepts such as solution directions and innovation paths, and that developments in these must be compared with developments in similar paths that have not benefited from the implemented policy. This is theoretically possible, because the essence of S/T policy is that at some point there must be a very clear focus within the change that the policy is trying to initiate or strengthen. In practice, however, it may be difficult to get to grips with this. When there are clear visions, such as roadmaps, that does not mean data on relevant investments and efforts can already be linked to the directions in those roadmaps. It is at least as important to be able to state which part of the data concerns directions that have not been selected;

- A third observation is that some experimental or quasi-experimental methods for evaluating S/T policies may be applicable generically to a very wide range of cases. Just as there are a number of standard techniques for single instruments, such as the use of (propensity score) matching or regression discontinuities, it is also conceivable that some approaches for S/T policies will end up in a general toolbox. This is obvious, for example, in network analysis, where there is often information in the policy records about which parties are involved in the policy and even who is collaborating. If S/T policies at actor level are intended to result in a stronger or different system,

this should be evidenced by changes in the size and/or structure of the networks involved. Although the most relevant indicators and types of networks can differ depending on the case, network analyses do therefore appear to merit a place in the toolbox. The same can be said of the approach described above to identify similar directions that are and are not supported by policy, in order to see whether there are differences (at high levels of development) in the degree of crowding-in. In summary, the call here is therefore for future evaluations of S/T policies to consider not only which experimental or quasi-experimental methods fit best, but also to look at the broader applicability of those experimental or quasi-experimental methods.

For qualitative methods, there is more freedom of choice in design compared to an experimental or quasi-experimental evaluation design that is severely limited by data availability. Which qualitative method is used depends greatly on the evaluator's choice as to which aspects of the policy he or she wants to focus on. In the case studies, a particular method was sometimes chosen that seemed most logical or promising in the light of the selected evaluation perspective and hypotheses. However, if different emphases had been applied in steps 1-7, the choice of qualitative methods could also have turned out differently.

9. Epilogue

This report seeks to explain what S/T policies are; why evaluating them is difficult; which 'perspectives' (and underlying analysis frameworks) can be chosen when evaluating the chain from intervention rationale to impact; and which empirical methods are likely to be the best to use.

Although scientific literature and insights from evaluation practice have been used as far as possible, it has been observed in various places that the evaluation of S/T policies is still in its infancy. The aim of this report is to take the first step in evaluating S/T policies, based on the advisory report of the Effect Measurement Expert Working Group. Future evaluations should show to what extent it is possible to subject all or parts of S/T policies to evaluations with a high causal standard of evidence. It should also become clear which concrete analysis frameworks are to be used or developed for each of the 'perspectives' on S/T policy evaluations referred to in this report. Since the evaluation of S/T policies is still at such an early stage, and because it will ultimately always require a customised approach, the recommendation is to conduct a review in a few years' time. Progress will undoubtedly have been made, enabling the content of this report to be refined. After all, as in the case of S/T policies themselves, experimentation is sometimes the only way to determine what really works. By also applying this adage to policy evaluations themselves, we can work towards establishing a tried-and-tested evaluation system.

Source material

Arnold, E. (2022). *Evaluating Missions*. Presentation for EC Research WP.

Arnold, E., Åström, T., Glass, C. and De Scalzi, M. (2018). *How should we evaluate complex programmes for innovation and socio-technical transitions?*

Åström, Arnold & Olsson (2022). Evaluating the 17 Strategic Innovation Programmes (SIPs). Technopolis Sweden

Bergek, A., Jacobsson, S., Carlsson, B., Lindmark, S., and Rickne, A. (2008). Analyzing the functional dynamics of technological innovation systems: A scheme of analysis. *Research policy*, *37*(3), 407-429.

Boon, W., & Edler, J. (2018). Demand, challenges, and innovation. Making sense of new trends in innovation policy. *Science and Public Policy*, *45*(4), 435-447.

Barbrook-Johnson, P., Castellani, B., Hills, D., Penn, A., & Gilbert, N. (2021). Policy evaluation for a complex world: Practical methods and reflections from the UK Centre for the Evaluation of Complexity across the Nexus. *Evaluation*, *27*(1), 4-17.

Dialogic (2017). Evaluatie SBIR.

Duflo, E. (2017). The economist as plumber. *American Economic Review*, *107*(5), 1-26.

Elzinga, R., Negro, S. O., Janssen, M. J., Wesseling, J. H., & Hekkert, M. P. (2020). Het Missiegedreven Innovatiesysteem: Uitbreiding 'Technologisch Innovatie Systeem'-raamwerk ter monitoring van de Circulaire Economie. Utrecht University.
Expertwerkgroep Effectmeting / Commissie Theeuwes (2012). Durf te meten.

Folkert, R., Verwoerd, L., & Verwest, F. (2018). Lerend evalueren: navigeren tussen verantwoorden en leren. *Beleidsonderzoek Online July*.

Foray, D., International Journal of Industrial Organization, 1997, The dynamic implications of increasing returns: Technological change and path dependent inefficiency, 15(6), 733-752.

Grillitsch, M., Hansen, T., Coenen, L., Miörner, J., & Moodysson, J. (2019). Innovation policy for system-wide transformation: The case of strategic innovation programmes (SIPs) in Sweden. *Research Policy*, *48*(4), 1048-1061.

Hekkert, M. P., Suurs, R. A., Negro, S. O., Kuhlmann, S., & Smits, R. E. (2007). Functions of innovation systems: A new approach for analysing technological change. *Technological forecasting and social change*, *74*(4), 413-432.

Hekkert, M. P., Janssen, M. J., Wesseling, J. H., & Negro, S. O. (2020). Mission-oriented innovation systems. *Environmental Innovation and Societal Transitions*, 34, 76-79.

Hünermund, P., & Czarnitzki, D. (2019). Estimating the causal effect of R&D subsidies in a pan-European program. *Research Policy*, *48*(1), 115-124.

Janssen, M.J. (2019). What bangs for your buck? Assessing the design and impact of Dutch transformative policy. *Technological Forecasting and Social Change, 138*, 78-94.

Janssen, M.J. (2022). Legitimation and effects of mission-oriented innovation policies: A spillover perspective. *Hacienda Pública Española / Review of Public Economics*.

Janssen, M.J., Bergek, A., Wesseling, J. (2022). Evaluating systemic innovation and transition programmes: Towards a culture of learning. *PLOS Sustainability and Transformations*.

Janssen, M.J., Bogers, M., & Wanzenböck, I. (2020). Do systemic innovation intermediaries broaden horizons? A proximity perspective on R&D partnership formation. *Industry and Innovation, 27*(6), 605-629.

Janssen, M.J. & Abbasiharofteh, M. (2022). Boundary spanning R&D collaboration: Key Enabling Technologies and Missions as alleviators of proximity effects? *Technological Forecasting and Social Change, 108,* 121689.

Janssen, M., Torrens, J., Wesseling, J., & Wanzenböck, I. (2021). The promises and premises of mission-oriented innovation policy: A reflection and ways forward. *Science and Public Policy*, 48 (3), 438-444.

Kroll, H. (2019). How to evaluate innovation strategies with a transformative ambition? A proposal for a structured, process-based approach. *Science and Public Policy*, 46 (5), 635-47.

Kuhlmann, S., and Rip, A. (2018). Next-Generation innovation policy and grand challenges. *Science and Public Policy, 45*(4), 448–454.

Larrue, P. (2021). The design and implementation of mission-oriented innovation policies: A new systemic policy approach to address societal challenges. Paris: OECD.

Loorbach, D., Frantzeskaki, N., & Avelino, F. (2017). Sustainability transitions research: transforming science and practice for societal change. *Annual Review of Environment and Resources*, *42*.

Lundvall, B. A. (1992). National systems of innovation: An analytical framework. *London: Pinter*.

Magro, E., and Wilson, J.R., (2019). Policy-mix evaluation: Governance challenges from new place-based innovation policies. *Research Policy 48,* 103612.

Ministry of Economic Affairs. (2017). Sturen in een verweven dynamiek: Perspectieven op complexiteit en oriëntaties voor beleid. The Hague: Ministry of Economic Affairs.

evaluatie van het Bedrijvenbeleid door het ministerie van EZK (2020)

Molas-Gallart, J., Boni, A., Giachi, S., & Schot, J. (2021). A formative approach to the evaluation of Transformative Innovation Policies. *Research Evaluation*, *30*(4), 431-442.

Patton, M. Q. (2011), Developmental evaluation: applying complexity concepts to enhance innovation and use. New York and London: The Guilford Press. ISBN: 978-1-60623-87.

PBL report “Grote opgaven in een beperkte ruimte” (April 2021),

Rodrik, D. (2008). *One economics, many recipes: globalization, institutions, and economic growth*. Princeton University Press.

Rogge, K. S., & Reichardt, K. (2016). Policy mixes for sustainability transitions: An extended concept and framework for analysis. *Research Policy*, *45*(8), 1620-1635.

Schot, J. and Steinmueller, W.E. (2018). Three frames for innovation policy: R&D, systems of innovation and transformative change. *Research Policy 47* (9), 1554-1567.

Scotchmer, S. (1991), Standing on the Shoulders of Giants: Cumulative Research and the Patent Law. *Journal of Economic Perspectives* 5 (1), 29–41.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., and Befani, B. (2012). Broadening the range of designs and methods for impact evaluations. *DFID Working Paper, 38*.

Turnheim, B., Berkhout, F., Geels, F., Hof, A., McMeekin, A., Nykvist, B., & van Vuuren, D. (2015). Evaluating sustainability transitions pathways: Bridging analytical approaches to address governance challenges. *Global Environmental Change*, *35*, 239-253.

Van Mierlo, B., Leeuwis, C., Smits, R., and Woolthuis, R. K. (2010). Learning towards system innovation: Evaluating a systemic instrument. *Technological Forecasting and Social Change, 77*(2), 318–334.

Van der Knaap, P., Pattyn, V. & Hanemaayer, D. (2020). Beleidsevaluatie in theorie en praktijk.

Van der Steen, M., Faber, A., Frankowski, A., & Norbruis, F. (2018). Opgavegericht evalueren. Beleidsevaluatie voor systeemverandering. The Hague: NSOB.

Verwoerd, L., Klaassen, P., van Veen, S. C., De Wildt-Liesveld, R., & Regeer, B. J. (2020). Combining the roles of evaluator and facilitator: Assessing societal impacts of transdisciplinary research while building capacities to improve its quality. *Environmental Science & Policy*, *103*, 32-40.

Weber, K.M. and Rohracher, H. (2012). Legitimizing research, technology and innovation policies for transformative change: combining insights from innovation systems and multi-level perspective in a comprehensive failures framework. *Research Policy, 41*, 1037–1047.

Wesseling, J.H., and Meijerhof, N. (2021). Developing and applying the Mission-oriented Innovation Systems (MIS) approach. DOI: 10.31235/osf.io/xwg4e

Wieczorek, A. J., & Hekkert, M. P. (2012). Systemic instruments for systemic innovation problems: A framework for policy makers and innovation scholars. *Science and public policy*, *39*(1), 74-87.

Wittmann, F., Roth, F., Hufnagl, M., Yorulmaz, M., Lindner, R. (November 2021). Second Mission Analysis Report of the Scientific Support Action to the German Hightech Strategy 2025 - Zooming in: Translating missions into policy instruments. Fraunhofer ISI. Karlsruhe.

Woolthuis, R. K., Lankhuizen, M., & Gilsing, V. (2005). A system failure framework for innovation policy design. *Technovation*, *25*(6), 609-619.