

12-6-2023

Navigating in Uncertainty

Developing an evaluation
approach for transformative
innovation policies and identifying
conditions for its
institutionalisation



Universiteit Utrecht

Copernicus Institute of Sustainable Development



**Ministerie van Infrastructuur
en Waterstaat**

Sander van de Wijngaert

Programme: Innovation Sciences
MSc Thesis

Utrecht University

Supervisor: Prof. dr. Koen Frenken
Second assessor: Dr. Joeri Wesseling

Ministry of Infrastructure and Water Management
Supervisors: Rik Braams & Henny de Jong
The Hague

Abstract

Innovation policies are increasingly aimed at addressing grand societal challenges. These Transformative Innovation Policies (TIPs) aim to fundamentally change sociotechnical systems such as energy, transport and food. Because TIPs are generally one of many interventions aiming for transformative change and their impacts appear a long way upstream from their implementation, they pose a substantial evaluation challenge. Current evaluation frameworks are mostly focused on single measures with clear predefined goals and a focus on policy outputs rather than outcomes or final impact, making them unable to capture specific transformative-related evaluation challenges. This study acknowledges the government's active role in steering transitions and set out to develop an evaluation approach suitable to the evaluation of TIPs and identify necessary conditions for the adoption of this approach by governments. In synthesising transition literature that attributes transition tasks to governments with policy evaluation literature, this research argues that these transition tasks should be adopted as policy objectives and therefore must be subject to evaluation. Moreover, it discusses that it is likely that tensions arise in policy evaluation practices as TIP evaluation asks for learning-based evaluations while policymakers are bound to traditional normative frameworks that favour accountability.

Four consecutive analytical steps were conducted. This research analysed whether and how transition tasks were evaluated in TIP evaluation whereafter these observations were compared with expert perspectives on current evaluation practices to propose an approach that suits the evaluation of transition tasks. An evaluation approach is proposed that evaluates whether conditions are created to successfully execute transition tasks. To do so, the evaluation approach suggests a reflexive process through stakeholder inclusion that is based on a flexible Theory of Change. Discussing challenges in evaluating transition tasks with policy evaluation experts led to the identification of barriers to the adoption and implementation of the proposed approach. Internal government structures are in place that hold civil servants accountable for their actions rather than providing room for learning. Through an illustrative case study, this research finds that current barriers to successfully executing transition tasks can be resolved when the reflexive approach is implemented early in the policy process. Moreover, the case study identified favourable conditions for learning. Several recommendations are made to legitimise learning within governments in an accountability-dominated evaluation culture.

Preface

This master's thesis was written for my graduation from the programme Innovation Sciences at Utrecht University. I had the valuable opportunity to enrich my thesis with an internship experience at the Dutch Ministry of Infrastructure and Water Management. The Ministry internship helped me understand how government organisations operate and how challenging it can be to change bureaucratic organisations from within. Through weekly team meetings, I had the opportunity to experience the work and effort my team made to change mindsets and make people more innovative. I encountered an incredible intrinsic drive from my colleagues to change themselves and others to make the Ministry a more transformative organisation. With this thesis, I hope to contribute to demonstrating why learning-based evaluations are necessary for transitions and to some extent open minds in seeing the value of these evaluations.

I would specifically like to thank Rik Braams and Henny de Jong for assisting with my research. Rik, thank you for challenging me at a high academic level and continuously asking me to substantiate my theoretical and methodological choices. This resulted in the layered approach of my thesis and the consciousness about every step I took. Moreover, you showed me the internal challenges civil servants face when they want to act transformative within the Ministry. Henny, thank you for guiding me through the complex world of Ministries, helping me in finding the right people and showing me how to be an entrepreneurial civil servant. Your great enthusiasm about how things 'were' possible instead of arguing how they were not, opened doors and positively affected my spirit. Last, I would like to thank my colleagues in the Innovation for Mobility team that specifically sparked my interest in the mobility transition we face as a society today. Mobility is so much more than only transport; it affects, amongst others, urban and special planning, infrastructure, energy supply and social inclusion. I hope this thesis is a first, and certainly not my last, contribution to how we can accelerate the mobility transition.

Enjoy the read,

Sander van de Wijngaert

June 12, 2023

Table of content

Abstract.....	1
Preface	2
1. Introduction.....	4
2. Literature	6
2.1. Three frames of innovation policy.....	6
2.2. Transition literature	7
2.3. Policy evaluation literature	11
2.4. Evaluating transformative innovation policies	14
3. Methods	16
3.1. The Dutch Ministry of Infrastructure and Water Management.....	17
3.2. Step 1: analysing transition tasks in TIP designs	17
3.3. Step 2: the evaluation of transition tasks in TIPs	18
3.4. Step 3: designing an evaluation approach and identifying tensions	18
3.5. Step 4: A case study of the Dutch Mobility as a Service Program	20
4. Results.....	22
4.1. Transition tasks in TIPs.....	22
4.2. A suitable approach for TIP evaluation	28
4.3. Tension in evaluation practices	30
5. Case study.....	31
5.1. The implications of the proposed approach.....	31
5.2. Favourable conditions for learning.....	33
6. Towards a learning-based evaluation framework.....	35
7. Conclusions	37
8. Discussion.....	38
8.1. Theoretical contributions.....	38
8.2. Limitations and suggestions for further research.....	39
References.....	41
Appendix A: Reviewed chapters of the annual budget of I&W	49
Appendix B: Codebook (Step 1 in methods)	49
Appendix C: Interviewee participants	50
Appendix D: Interview guide.....	50
Appendix E: Coding tree (Step 3 in methods)	52
Appendix F: Coding tree (Step 4 in methods)	53
Appendix G: description of abbreviations in Table 5	53

1. Introduction

In recent years, increasing efforts are made devoting policy to solving grand societal challenges, illustrated by the adoption of the Paris Climate Agreement and the Sustainable Development Goals (Kuhlmann & Rip, 2018). Facing key societal challenges requires a fundamental shift towards sustainable sociotechnical systems (Schot & Steinmueller, 2018; Weber & Rohracher, 2012), where governments play an essential role in facilitating and steering the direction of these transitions (Borrás & Edler, 2020; Hekkert et al., 2020). Transformative change refers to a fundamental and systemic shift in socio-technical systems such as energy, transport and food, and includes changes in behaviour, culture and governing institutions (Rohracher et al., 2023). Transformative change is at the heart of two perspectives of innovation policy: transition-oriented innovation policies, which draw on innovation and transition studies, and mission-oriented innovation policies, which address well-defined societal objectives in a defined timeframe (Diercks et al., 2019). These works of literature have recently started to move in the same conceptual direction as they come together in a new innovation policy paradigm: Transformative Innovation Policy (TIP). TIP combines insights from sustainable transition literature (Schot & Steinmueller, 2018) and a broader understanding of the innovation process and a societal policy agenda (Diercks et al., 2019). Transformative change calls for a new discussion on directionality which implies that innovation policy should go beyond merely technological innovation and economic growth. These policies should be aimed at addressing societal challenges instead (Diercks et al., 2019; Schot & Steinmueller, 2018; Weber & Rohracher, 2012).

As TIP diffuses and gains legitimacy, there is a need for suitable evaluation methods to assess the impact of policies aiming for transformative change (Janssen et al., 2022; Mazzucato, 2016). However, the impacts of TIPs are often visible a long way upstream from their implementation which makes it difficult to evaluate policy impacts in a short time frame as done with traditional evaluation frameworks (Amanatidou et al., 2014; Haddad et al., 2022; Janssen, 2019; Molas-Gallart et al., 2021). Moreover, TIPs are generally only one of many interventions aiming for transformative change making them part of complex policy mixes. This makes it much harder causally attribute certain outcomes to a specific intervention (Janssen, 2019; Kivimaa & Kern, 2016; Molas-Gallart et al., 2021). In these situations, the outcomes of an intervention can be no more than a contribution to the eventual system changes policy interventions aim for. As TIPs aim for systemic change, the assessment of policies should measure the impact on selected challenges, rather than the innovation outcomes of policy in general, which existing evaluation frameworks are unable to capture (Haddad et al., 2022).

Literature highlighting new rationales for policy intervention implies that new aims and tasks for policy ask for adapted evaluation practices (Amanatidou et al., 2014). Current evaluation frameworks are mostly focused on single measures with clear predefined goals and are based on the linear view of the innovation process (Amanatidou et al., 2014) with a focus on policy outputs rather than outcomes or final impact (Lee, 2015). Multiple scholars adapt existing frameworks which mainly build on previous approaches to evaluating systemic innovation policies (e.g. Janssen, 2019; Kivimaa & Kern, 2016; van Mierlo et al., 2010). These frameworks are considered useful for policy analysis but often define certain technological solution. Transformative policies typically deal with 'wicked problems' (Rittel & Webber, 1973) for which no straightforward solution exists. Therefore, these frameworks are not fully equipped to address the evaluation of specific transformative-related directionality challenges (Haddad, 2021).

Some studies suggest evaluation approaches for TIPs that adhere to TIP-specific challenges (Molas-Gallart et al., 2021; Rohracher et al., 2023). However, these evaluation approaches are not specifically directed at governments, who have an essential role in implementing and evaluating TIPs to address societal challenges (Borrás & Edler, 2020; Hekkert et al., 2020). Moreover, little research has been done in identifying specific government challenges in implementing and

adopting these new evaluation approaches. Considering that the government plays an essential role in providing funding, setting goals and targets, and establishing frameworks for evaluation (Mazzucato, 2016), the aim of this research is twofold: 1) to develop an evaluation approach specifically directed at governments that is suitable to the evaluation of TIPs and; 2) to identify necessary conditions for the adoption of TIP evaluation approaches by governments. Based on these objectives, the following research question is formulated: *how does the shift towards TIP as the third frame of innovation policy create challenges in policy evaluation practices and what new evaluation principles and conditions can we develop to overcome this?* This research combines insights from two strands of literature – transition literature and policy evaluation literature – together with empirical observations and expert perspectives, to address challenges and uncover tensions in evaluating TIPs.

To answer this question, this research is structured as follows. The next section elaborates on transition literature and policy evaluation literature in the context of the development of TIP. Section 3 describes the used methods, where four consecutive steps were taken to answer the research question. Subsequently, section 4 proposes an evaluation approach suitable for TIP, and presents the identified tensions that hamper the institutionalisation of learning-oriented evaluation approaches. These outcomes are illustrated in a case study in Section 5. Section 6 reflects on what is necessary to move towards a learning-based evaluation framework. Finally, sections 7 and 8 conclude this research by summarizing and discussing the main finding and contributions of this study.

2. Literature

This chapter provides the literature body of this research and highlights the development of different frames of innovation policy with TIP as the newest premise. Particular attention is paid to the role of governments in transitions by discussing five transition tasks attributed to governments that should be subject to evaluation. Furthermore, it discusses the challenges for evaluation and a fundamental shift in the role of policy evaluation parallel to the shift towards TIP.

2.1. Three frames of innovation policy

Schot and Steinmueller (2018) discuss the existence of three frames in innovation policy. The first frame which emerged during the 20th century is based on a linear understanding of innovation: innovation emerges from a process of knowledge creation through basic and applied research. This linear model often relies on supply-push market mechanisms for R&D, and market failures and externalities formed the rationale for policy intervention (Schot & Steinmueller, 2018). Further exploitation of knowledge into the application of technologies created welfare and growth. Policy objectives therefore focused on R&D inputs and how these shaped the knowledge system that led to innovation. This frame of innovation policy can be seen as a predominantly narrow understanding of the innovation process (Diercks et al., 2019). A strong belief in this top-down control of science was backed by scientific endeavours around nuclear energy and space programmes (Chiang, 1991; Ergas, 1986).

In the last decades of the 20th century, the linear model of innovation was seen as overly simplistic and was criticized for failing to provide indirect broader structural support and mechanisms for the diffusion of innovation (Chiang, 1991; Ergas, 1986). This gave rise to the second frame for innovation policy, which was based on systems of innovation that stressed the progression of innovation and growth through knowledge development. The implementation of technological development is far from automatic and does not move in a single direction. Innovation and growth were contingent upon a variety of institutional factors and the interaction between innovation system participants. Addressing the limitations of the first policy frame, a more holistic view formed the basis for innovation systems policy, consisting of different and yet complementary ways of framing innovation through national (Lundvall, 1992), regional (Cooke et al., 1997), sectoral (Malerba, 2002) or technological innovation systems (Carlsson & Stankiewicz, 1991). This provided policymakers with a framework to generate a set of policy recommendations by applying systems of innovation heuristics (Diercks et al., 2019). The rationale for innovation policy was based on systemic failures, moving beyond R&D investment to institutional conditions that can promote innovation. Within this frame, interactions among various actors in the innovation systems can be defined and operationalized by policy objectives focussing on the scope, scale and quality of these interactions.

Research on innovation systems is mainly directed at optimizing innovation systems to maximise economic policy objectives. But since the mid-2000s, a shift from an economic policy agenda towards a social policy agenda can be noticed, spurred on by the urgency of societal challenges (Kallerud et al., 2013). This gave rise to the third frame labelled as 'Transformative Innovation Policy' (Diercks et al., 2019). TIP stresses that innovation policy must not only optimize innovation systems to improve economic growth but also address societal and environmental challenges by transforming sociotechnical systems (Daimer et al., 2012; Schlaile et al., 2017; Weber & Rohracher, 2012). A key concern is the directionality perspective, which investigates social and environmental drivers, desirable policy directions and directions for innovation while blocking undesirable ones. To address directionality, TIP needs to incorporate deep learning and reflexive processes to question and reframe underlying assumptions about desirable directions as these directions cannot always be known from the start (Molas-Gallart et al., 2021). However, following this third frame does not mean that the first two frames should be abandoned. The TIP paradigm

can be seen as layered upon, but not fully replacing, the earlier policy paradigms of science and technology policy and innovation systems policy, as investing in R&D, system interactions, and learning among actors is still deemed relevant (Diercks et al., 2019). The third frame calls for a reorientation of policy in the first two frames towards transformation, by, for example, focusing R&D investments on Sustainable Development Goals.

2.2. Transition literature

Literature on sustainable transitions emerged to study the transformation of sociotechnical systems towards more sustainable environments (Kivimaa & Kern, 2016). These studies go beyond focusing solely on technological innovation but include system dynamics such as infrastructures, institutions, industrial sectors and the behaviour of users. Transition studies take a systemic perspective with an analytical focus on supporting innovation systems by drawing attention to novelty and destabilizing existing structures (Köhler et al., 2019). Salient views in the transition field are the multi-level perspective (MLP), strategic niche management (SNM) technological innovation systems (TIS), and transition management (TM). These perspectives on system change articulate directionality for TIPs to foster system change, summarized in Table 1. This section describes these transition studies and discusses the role of the government in directing TIPs.

2.2.1. *Multi-level perspective*

The MLP argues that transitions emerge through dynamic processes within and between three system levels: 1) niches, which are protective spaces that nurture new innovations; 2) sociotechnical regimes, which represent the institutional structure of existing systems that favour stability and incremental change; and 3) exogenous landscape, which refers to the external endogenous processes that influence the regime (Geels, 2002; Smith et al., 2010). The MLP indicates that top-down landscape pressures and bottom-up developments of several niches create destabilization of incumbent regime structures. This offers opportunities for emerging niches to break through by interactions between the niche and regime level and eventually overthrow the incumbent regime.

Policy interventions derived from MLP literature should mainly focus on the support of niche developments (e.g., targeted R&D funding and creating visions) and the destabilization of incumbent regimes (e.g., taxes and removing subsidies) (Kanger et al., 2020). The standard rationale for policy intervention in innovation activities is based on the arguments of market failures (Arrow, 1962) and system failures (Woolthuis et al., 2005). Market and system failures address structural deficits in innovation systems, but do not fully align with the transformative processes identified in the multi-level perspective (Weber & Rohracher, 2012). While still highly relevant to justify policy intervention, from a transformative change perspective, these failures arguments are too restrictive and leave out important instances of failure. Weber & Rohracher (2012) therefore propose additional 'transformative failures' that address requirements for goal-oriented transformative change; directionality, demand articulation, policy coordination and reflectivity.

2.2.2. *Strategic niche management*

Connected to the MLP and developing simultaneously, SNM is a framework that is used for analysing the emergence of radically new innovations (Geels & Raven, 2006; Rip & Kemp, 1998; Schot & Geels, 2008). SNM scholars argue that radical innovation emerges in protected spaces that are shielded from market pressures, similar to the niches described in MLP literature. Through sequences of experiments and demonstrations, and interactions between learning processes, social networks, and visions and expectations, niche innovations can develop (Kemp et al., 1998) and can generate innovation trajectories (Geels & Raven, 2006).

Literature on SNM argues that policy intervention should focus on protecting niches and the alignment of niches with regime structures. Central directions for policy are to stimulate learning through the inclusion of a variety of actors, create shared visions, network development by including incumbents and outsiders, the protection of niches, and niche-regime interaction through stimulating niche innovations when alignment opportunities arise (Schot & Geels, 2008). Directionality towards a desired system should be stimulated through regulation by setting ambitious targets and penalties for niche players that are enforced when targets are not met (Caniëls & Romijn, 2008).

2.2.3. *Technological innovation systems*

A related theoretical trajectory is that of the TIS, which is developed to study the emergence of new technologies and the formation of innovation systems around them (Bergek et al., 2008; Hekkert et al., 2007; Negro et al., 2008). Technological development and diffusion are the results of a positive fulfilment of system functions influencing the innovation system around a particular technology. Through cumulative causation, different functions strengthen each other, leading to self-reinforcing motors of innovation, allowing a TIS to develop (Suurs & Hekkert, 2009).

The TIS approach shows that instead of policy targeting the supply-side (R&D programmes) or the demand-side (market creation), policy should contribute to the formation of technological innovation systems and thereby enhance the chance of success for new technologies. The system functions provide directionality for policy and suggests that policy should address lacking system functions (Suurs & Hekkert, 2009). To determine suitable policy targets for certain challenges it should be investigated what blocking mechanisms are in place that hamper the fulfilment of certain system functions (Bergek et al., 2008).

2.2.4. *Transition management*

TM is a policy-oriented, prescriptive framework, which proposes four sequential activities that help policymakers shape transitions (Loorbach, 2010; Rotmans et al., 2001); 1) strategic activities aimed at developing a vision and the identification of transition pathways; 2) tactical activities, which build agendas and support coalitions for more specific transition pathways; 3) operational activities, including activities like demonstration projects, experiments and activities aimed at learning-by-doing; 4) reflective activities, leading to adjustments of visions, by the monitoring and evaluation of projects and progress.

Table 1: policy articulation from transition studies

Transition literature	Policy articulation
MLP	Create directionality Articulate demand Coordinate policies Provide spaces for experimentation and learning
SNM	Stimulate learning through the inclusion of a variety of actors Create a shared vision Network development by including incumbents and outsiders Protection of niches Niche-regime interaction through stimulating niche innovations when alignment opportunities arise
TIS	Address lacking system functions by identifying blocking mechanisms Promote motors of innovation
TM	Create a transition arena Develop visions that guide the formulation of policy measures Back-casting of goals to identify alternative transitions pathways Explore transition paths through practical experiments Continuously monitor and evaluate experiments to revise the guiding visions

Transition management departs from a set of persistent problems that cannot be solved by conventional policy approaches in the context of incumbent structures. In TM, the approach to policy design comprises five main components (Loorbach, 2007). A transition arena needs to be established, i.e., a platform for transition-oriented interactions amongst actors. These arenas stimulate creative interaction, the exchange of knowledge, learning and discussion among participants. Within the transition arena, general policy goals should be translated into visions that guide the formulation of policy measures. These visions can be realized through back-casting, which generates alternative transition pathways that link the visions with the present (Quist, 2007). These transition paths can be explored through practical experiments. In turn, these experiments can adjust developed visions and pathways. The overall process as well as the specific experiments should be continuously monitored and evaluated to be able to revise the guiding visions.

2.2.5. *Transformative government*

From the articulated policy recommendations in transition literature, tasks have originated that call for governments to engage with the deeply rooted societal problems that require societal transitions (Bergek et al., 2015; Kivimaa & Kern, 2016). Braams et al. (2021) reviewed 100 articles on transition literature that attributed transition tasks to governments to identify these transition tasks. This resulted in five clusters of transition tasks (see Table 2 for task descriptions and specific transition tasks): 1) give direction; 2) support governance; 3) support the new; 4) destabilize the unsustainable and; 5) develop internal capabilities and structures.

The implementation of TIPs to address societal challenges and transform systems require governments to have extensive knowledge and organisational capacity to engage in the systems they are trying to change (Kattel & Mazzucato, 2018). In a study on public administration traditions that provide legitimacy to governments, Braams et al. (2021) highlight that existing traditions are mostly unsuitable to empower civil servants in fulfilling these transition tasks. They propose a new tradition – transformative government – that legitimizes the execution of transition tasks, i.e., “a government that understands, accepts and executes transition tasks ... It synthesizes notions of system change with an understanding of administrative processes, legitimacy, and democracy to enable a legitimized pursuit of transition tasks” (Braams et al., 2021, p. 200). It builds on the idea that governments should take on transition tasks that are not sufficiently fulfilled by society and should design systems in a way that aligns with social and environmental challenges for the entire duration of the transition. The rationale for policy intervention is based on the idea that the government is the guardian of interests that are not sufficiently represented by politics, the market or society (Braams et al., 2021).

This research adopts the concept of *transformative outcomes* that “offer guidance about the transformative change that we need to trigger and thus help profile a policy in terms of its transformative potential” (Molas-Gallart et al., 2021, p. 435). Schot et al. (2019) and Ghosh et al. (2021) draw these transformative outcomes from three core transition processes from MLP literature: building and nurturing niches, expanding and mainstreaming niches into the wider world, and opening up and unlocking regimes. In their paper on evaluation methods for TIP, Molas-Gallart et al. (2021) argue that an evaluation approach needs to focus on these transformative outcomes to assess the degree to which interventions are progressing towards the achievement of long-term systemic goals. Although the transformative outcomes resonate with *support governance* and *support the new*, the transition tasks *give direction*, *destabilize the unsustainable* and *develop internal capabilities and structures* are overlooked. Yet, these tasks are essential to accelerate a transition.

Table 2: attributed transition tasks to governments by transition literature (Braams et al., 2021).

Categories of transition tasks	Task description	Specific transition tasks
<i>Give direction</i>	Governments should give directionality through the articulation of demand and shared visions. Leadership should be taken in establishing ambitious policy objectives and policy strategies should steer innovation towards societal needs.	Articulate the direction
		Construct policy strategies to direct
		Reconfigure the market
		Direct through enforced regulations
<i>Support governance</i>	The government should open up the process of transition for multiple stakeholders and encourage others to participate through collective action. Additionally, governments should develop and maintain network relations and are responsible for collective outcomes within these networks.	Activate actors
		Guiding organizational arrangements
		Goals achieving strategies
<i>Support the new</i>	Governments should support, fund and engage with new bottom-up developments. Interventions should stimulate niche development as they could lead to new configurations of incumbent regimes over time (Kivimaa & Kern, 2016). As such, the government must engage with, facilitate and fund new developments.	Engage in entrepreneurial experiments
		Establish market formation
		Price-performance improvements and resource mobilization
		Help new developments develop and diffuse
<i>Destabilize the unsustainable</i>	It recognizes that governments should destabilize undesired sociotechnical regimes (Loorback, 2007; Rotmans et al., 2001). This involves the pressuring and phasing out of specific regime processes to create opportunities for niche innovation to align with regimes.	Control policies and make significant changes in regime rules
		Reduce support for dominant regime technologies
<i>Develop internal capabilities and structures</i>	Governments should internally focus on developing capabilities and structures to facilitate external tasks. Through developing in-house skills and structures, governments can enhance their ability to participate in promoting and directing social transitions.	Rethink own role in a transition
		Development of new competencies
		Monitor and evaluate
		Establish mechanisms for policy coordination

This research acknowledges that the MLP processes are essential for transitions, but the transformative outcomes only build on MLP literature and are not explicitly directed at governments. They neglect several government tasks ascribed from transition literature. Because the execution of tasks is attributed to governments (Braams et al., 2021) and governments should intervene with innovation policy when society fails to execute the tasks (Weber & Rohrer, 2012), one can argue that empowering civil servants to execute transition tasks should become a policy objective in itself. Following Braams et al. (2021), a transformative government is able to fulfil all five tasks and therefore the tasks are desirable policy outcomes. This research proposes

that the five transition tasks should be adopted as policy outcomes and should therefore be subject to evaluation.

2.3. Policy evaluation literature

Policy instruments, measures or programmes are aimed at achieving desired goals. This raises the question of whether those goals are indeed achieved by the incentivised policy. Policy evaluation aims to assess new and existing policies and inform policymakers about the effectiveness and efficiency of the policy. The following section elaborates on the theory behind policy evaluation.

2.3.1. Seven functions of policy evaluation

Before evaluating policy, one should think about what the goal of a certain evaluation is and what one wants to accomplish by evaluating policy. Van der Knaap et al. (2020) identify seven functions of policy evaluation. Each function represents a specific goal of evaluation and to what kind of knowledge it contributes.

1. *Informing*: it is impossible to assess new or existing policies or make well-informed decisions without sufficient information. This function of policy evaluation informs policymakers, people who make policy decisions or people affected by the policy about all relevant knowledge on the policy measure. It addresses what is known about certain policies and what additional information needs are.
2. *Judging*: the question often asked with policy is whether the government is doing well in achieving policy goals and whether its actions are properly executed. Judging is the normative function of policy evaluation where one is forming an opinion on the 'value' of activities, processes, products or effects related to policy. Judging the value of policies can eventually lead to decision-making about policies.
3. *Learning and adjusting*: learning is one of the central functions of evaluation because improving policy requires learning about what could be improved (Wildavsky, 1979). Learning is based on feedback mechanisms induced by monitoring and can be first or second-order learning (Argyris & Schön, 1997). In first-order learning, an error is detected and corrected which allows the addressing of present objectives. In second-order learning, an error is detected and corrected in ways that involve modification of an organization's underlying norms, policies and objectives.
4. *Accountability and justification*: before executing policy and spending public money, policymakers need to get approval from a governmental body. Ex-ante or ex-post evaluation can justify policy measures by informing about the legitimacy of a certain policy (Keulen, 2020). The most important aspect of policy evaluation in accountability is contributing to controlling the public exercise of power by publicly reporting successes and failures of policy.
5. *Rationalising*: theory development and testing informs political decision-makers about causal relations of problems and solutions (Hoogerwerf, 1984). Several policy questions are central: what is the problem? what causes the problem? and who suffers from the problem? Interventions should preferably be formulated as SMART-C: specific, measurable, acceptable, realistic, time-bounded, and consistent in definition and measurement methods. To assess policy effectiveness, criteria can be used to compare different interventions, including a null option.
6. *Discussion*: evaluation reports and evaluation methods can open discussion on executed policy or on how policy should be executed. Discussions between actors contribute to the exchange of information and perspectives which gives input on what practical impact policy has (Van der Knaap, 1995). Evaluators that favour the dialogue between actors warn against traditional top-down approaches that are more static and hold on to a chosen policy goal. They believe that a top-down approach could hamper policy execution by the inability to accommodate different realities and an over-belief in the scientific paradigm.

7. *Politicizing*: policy evaluation can be strategically employed, for example, by taking more time before making important decisions. However, the need for more evaluation research can also be exploited. Evaluations can be used to postpone or manipulate decision-making, frame political opponents and avoid responsibility (Banner, 1974).

2.3.2. *Three types of policy evaluation*

This section describes three types of policy evaluation, developed in reaction to the changing role of government and policy in society, that originate from philosophical movements; rationalism, constructivism and contextualism (van der Knaap et al., 2020).

2.3.2.1. *Type I: Rational-analytical policy evaluation*

The rational-analytical type of policy evaluation, developed in the 1960s, emphasises the goal of scientifically measuring the efficiency and effectiveness of policies. The government is the central actor that should make well-balanced decisions (Abma & in 't Veld, 2001). The achievement of policy goals is central and policy should be based on a social cost-benefit analysis of policy alternatives, policy instruments are chosen based on proven effectiveness and policy execution is done based on cost-effectiveness (van der Knaap et al., 2020; van Mierlo et al., 2010). According to this policy theory, a hypothesis is developed where policy instruments are linked to policy targets. A condition for determining the effectiveness and efficiency of policy is that clear policy goals must be formulated, preferably quantitative and with final terms.

Although rational-analytical evaluation often emphasizes the determination of causal relations between policy and intended effect, evaluations can also investigate possible explanations of disappointing or successful results. Type I evaluation focuses mostly on what factually happened during policy execution, e.g., if predefined targets were met and whether this was according to the agreed planning and agreed-on quality (van der Knaap et al., 2020). This can uncover problems that can be corrected through first-order or second-order learning. However, these findings uncover if a policy was well-executed, but do not explain why causal relations or policy theory fails. According to the rational-analytical perspective, causal relations are preferably developed in the policy preparation phase (*ex-ante*).

2.3.2.2. *Type II: Constructivist policy evaluation*

The constructivist type of policy evaluation emerged during the 1980s with the prominent motive to explain undesired policy results from a broader perspective than only the shortcomings in the execution of policy (van der Knaap et al., 2020). Where rationalists assume that knowledge about policy effects and efficiency are the product of quantitative methods, constructivists believe that policy goals and effects bring value when stakeholder perspectives are included (Guba & Lincoln, 1989). The key concept in Type II evaluation is social plurality, meaning that researchers should incorporate the variety of visions of stakeholders. According to constructivists, a centrally determined and 'simplistic' policy theory cannot be the start for policy evaluation. Subjective judgements about policy are not considered problems that should be solved with research but are an enrichment that should be reconstructed. Plurality, responsiveness and inclusivity are central instead of efficiency and effectiveness (Abma, 1996).

The constructivist approach aims to utilize the quality of policy and policy evaluation by including the perceptions and experiences of involved actors. Through cognitive learning, every individual or organisation has its own truths and beliefs of what problems and solutions are deemed desirable, and what desirable goals and suitable instruments are. Constructivist policy evaluation is an iterative process where stakeholders come to new insights about problems and policy options through discussion and argumentation (van der Knaap et al., 2020; van Mierlo et al., 2010). This will increase the involvement of stakeholders in the evaluation, which will, in turn, increase the chance of the application of evaluation outcomes.

2.3.2.3. Type III: Contextual-realist policy evaluation

Contextual-realistic policy evaluation, or realist evaluation, emerged in the 2000s and combines elements from type I and type II policy evaluation. Type III policy evaluation emphasises the specific context of policy and executions to identify successful mechanisms (Pawson & Tilley, 1997; van der Knaap et al., 2020). As well as the rational-analytical approach, the contextual-realist approach is theory-based where hypotheses about policy effects are central. However, in Type III policy evaluation the question is not ‘Does policy work?’ or ‘Is the intervention effective?’, but ‘What policy works, through what mechanisms, for whom and under which circumstances?’ (Pawson & Tilley, 1997). Policy effects should be seen as the result of all interaction in and between systems and not solely the result of policy (Abma & in ’t Veld, 2001). In evaluation literature, contextual-realistic policy evaluation is described with the Intervention, Context, Mechanism, Outcome (ICMO) model. It is focused on uncovering mechanisms (M) that are initiated by the intervention (I) in a specific context (C) that lead to certain outcomes (O) (Pawson & Tilley, 1997).

Type I evaluation compares a situation with a policy intervention to a counterfactual. If one situation shows a significantly different result than the other, a rational analytic concludes that the difference is the result of policy intervention. However, what exactly explains the difference is not uncovered. This is most problematic when no effect is found: the hypothesis is rejected but no explanation is given. Type II and type III approaches focus on the explanations by uncovering underlying factors of why a policy intervention is not effective (Pattyn & Verweij, 2014). A summary of the three evaluation types is given in Table 3.

Table 3: functional focus, advantages and disadvantages of policy evaluation types (van der Knaap et al., 2020)

Evaluation type	Functional focus	Advantages	Disadvantages
Rational-analytical	Rationalising Informing Learning Justification	Clear research design: does policy work? Emphasis on effectiveness	Simplistic and neglects social pluralism High methodological standards for experimental research The complexity of policy and its context complicates application Does not explain success or failure
Constructivist	Rating Learning Discussion	Focus on social pluralism Open to opinions of other actors than the central actor Higher chance of application of evaluations	Alleged noncommittal Conclusions strongly depend on stakeholder perspectives
Contextual-realist	Rating Learning	Close to practical research Emphasis on explanations of working mechanisms in a specific context	Complex argumentation for mechanisms Time and context lead to expirable conclusions Conclusions can be seen as relativistic

2.3.3. Fundamental shift in the role of TIP evaluation

The change in policy goals and legitimisation towards addressing societal challenges constitutes a fundamental shift in the role of innovation policy and its evaluation (Schot & Steinmueller, 2018). In more simplistic policy situations, the main purpose of the evaluation is likely to be accountability and justification (Magro & Wilson, 2019). This led to more traditional evaluation approaches (as described by Guba and Lincon (1989) as first and second generations of evaluation

practices) which focus on evaluators assessing policies based on policy impact, effectivity and effectiveness. However, TIP impact, effectivity and effectiveness cannot easily be captured by more traditional policy evaluations. A major challenge for TIP evaluation is that system transformation often asks for a 'policy mix' that aims to find the most optimal combination of interactions and mutual support of different instruments, instead of finding one single best policy instrument with the most optimal result (Haddad et al., 2022; Kanger et al., 2020). This makes the distinction between individual policy effects and the effects of multiple transformative programs as a whole a difficult task. Moreover, TIPs are subject to complex interactions and long timescales at which societal challenges unfold. This complicates the accountability and justification of public policies through summative (ex-ante) evaluation as direct effects and attribution of policies to system change are difficult to assess.

Recent literature (e.g. Kattel & Mazzucato, 2018; Kuhlmann & Rip, 2018; Molas-Gallart et al., 2021; Rohrer et al., 2023) therefore pleads for more formative (ex-durante) policy evaluations that evaluate the progress towards a certain goal and centralizes the learning about the functioning of policies. In contrast to summative evaluations, a formative approach is a style of evaluation conducted with the participation of stakeholders. Others argue that policy evaluations should emphasise reflexivity of the policy process, i.e. a process that challenges and change presumptions, current practices, and underlying institutions of all stakeholders (Grin et al., 2004; van Mierlo et al., 2010; Voß & Kemp, 2015). The continuous questioning of dominant values and institutions allows for rapid policy adaptation aiming to improve policy interventions while being implemented (Janssen et al., 2022; Molas-Gallart et al., 2021). New evaluation frameworks for TIP should centralize the role of learning how to contribute to system innovation by putting prevailing values and institutional settings up for discussion. However, it remains challenging to facilitate policy learning in practice due to its complexity, potential to conflict with accountability and often lacks priority from involved stakeholders (Amanatidou et al., 2014; Magro & Wilson, 2019).

2.4. Evaluating transformative innovation policies

In the mid-20th century, a new management reform approach, known as New Public Management (NPM), arose as a response to the challenges faced by public organisations which were seen as bureaucratic, unresponsive and inefficient (Behn, 1998). NPM places a strong emphasis on using business-like management strategies to improve productivity, effectiveness and accountability. Policy evaluations have been a critical component of public policy-making and allowed policymakers to assess the effectiveness and efficiency of their actions and identify areas for improvement. Following the rise of NPM, policy evaluation has responded to the need for improved accountability for policymakers (Norris & Kushner, 2007). Establishing democratic and political accountability is the primary function of policy evaluation as it contributes to trust in governments by citizens and provides evidence regarding policy performance (OECD, 2020). It sets out to make governments more efficient and effective with one big advantage as a result: it holds a government accountable. A central tenet of NPM is its focus on performance measurement and accountability, which require the use of quantitative data and data analyses to evaluate policies and programs. As part of a broader shift to NPM, the rational-analytical evaluation approach has been widely adopted by governments. To assess the effectiveness and effectivity of public organisations, the rational-analytical approach allows policymakers to make evidence-based decisions and prioritize policies that deliver the greatest bang for their buck. Policy evaluation in the NPM perspective, therefore, holds policymakers accountable for the expenditure of public resources.

However, as discussed in section 2.2.3., academic scholars argue that learning should be a more dominant function of policy evaluation to be able to capture the uncertainty and complexity of TIPs. Given the principles of NPM based on justification and accountability, it is reasonable to

expect tensions in current evaluation practices. Policymakers and evaluators may favour rational-analytical evaluation approaches as they form a rationale for evaluation according to NPM, while scholars are critical of the rational-analytical approach as it fails to capture the complexity of TIPs. Reinertsen et al. (2022) demonstrate that evaluations focused on justification and accountability are often at the expense of learning about the policy process. Stakeholders may prioritize one over the other depending on their interests and priorities (Reinertsen et al., 2022; Weiss, 1998). Weiss (1998) notes that policymakers may be sceptical of evaluations that do not produce positive results or fail to demonstrate the effectiveness of their policies and may be more likely to use evaluations that support their policy positions or confirm their prior beliefs. She suggests that the focus on accountability and justification can sometimes result in a narrow or limited approach to evaluation, where the emphasis is on measuring inputs, outputs, and outcomes rather than on understanding the complex and dynamic processes that underlie policy interventions. The shift towards a new frame of TIP evokes a shift towards learning-oriented policy evaluation, which is often at tension with more 'traditional' policy evaluations focussed on accountability and justification.

In sum, this literature body discusses the debate in evaluating TIPs. It describes that a new frame of innovation policy is on the rise which aims at addressing societal challenges rather than stimulating economic growth. From transition literature, transition tasks emerge that are specifically attributed to governments and therefore should become policy objectives of TIPs. These policies, however, come with complex policy interactions which make it difficult to assess direct policy effect. Therefore, scholars argue for more learning-oriented evaluation approaches. From a historical perspective, governments must adhere to NPM principles that favour accountability rather than learning. This creates potential tensions in the field of policy evaluation that this research aims to uncover.

3. Methods

This research set out to develop an evaluation approach that is suitable for the evaluation of TIPs and analysed tensions in evaluation practices to establish necessary conditions for the institutionalization of TIP evaluation approaches. In doing so, this research conducted four consecutive steps. It analysed the presence of transition tasks in current TIPs and how these tasks were evaluated in TIP evaluations. This provided insights into the adoption of transition tasks in current policy practices and whether these tasks were subject to evaluation. This was followed by interviews to uncover problems and difficulties in evaluating the transition tasks using existing evaluation frameworks. Through discussing how the transition tasks could be included in evaluation practices, an evaluation approach suitable to the evaluation of TIPs was proposed. Moreover, the interviews allowed to discover why the shift towards learning, as a function of policy evaluation, remains difficult in the field of policy evaluation. A case study was conducted by examining a policy program which shows the implications of the proposed approach and analyses favourable conditions for adopting learning a policy goal. Figure 1 provides a schematic overview of the conducted steps.

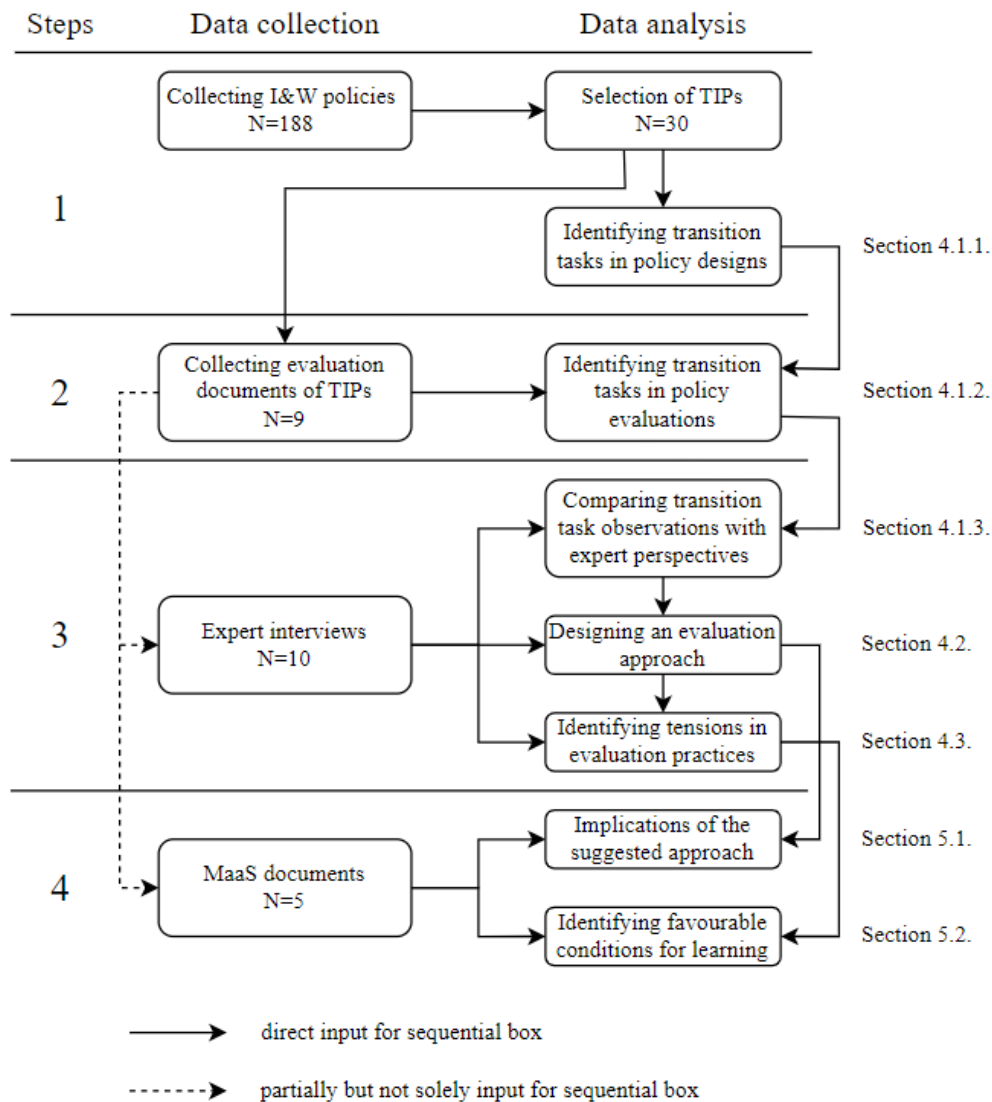


Figure 1: Flow diagram of methodological steps

3.1. The Dutch Ministry of Infrastructure and Water Management

A system aiming for transformative change is the Dutch mobility system. The Dutch Ministry of Infrastructure and Water Management (I&W) is working on a 'green and smart mobility system'¹ that is "focused on sustainable behavioural change and conscious mobility behaviour of citizens and companies." (I&W, 2021). In doing so, I&W designs policies that resonate with TIP characteristics in various ways. First, it incorporates societal challenges as its policy goals such as quality of life, safety and environment (Snellen et al., 2021). Second, the policy approaches that aim to address these societal challenges go beyond traditional innovation policy. By aiming for broad prosperity and not solely accessibility and growth, I&W develops policy according to the third frame of innovation policy. Last, I&W intends to transform the mobility system by aiming to pressure the current mobility regime and create a new one (I&W, 2016). This shift makes the Ministry of I&W an appealing place to study transformative innovation policies and the institutional context for implementing suitable approaches.

3.2. Step 1: analysing transition tasks in TIP designs

3.2.1. Data collection

The first step consisted of collecting transformative policies designed and conducted by I&W. The explanatory document of the policy budget 2022² of I&W was reviewed to identify current policies in the Dutch mobility domain. The policy agenda and multiple policy articles (Appendix A) of the explanatory document were studied to identify policy initiatives that are conducted by I&W. Initiatives from the Ministry were collected in a database which led to a collection of 188 initiatives across various mobility directorates. Internet search engines and Dutch government websites were consulted to obtain policy designs of the 188 initiatives describing their policy goal. The policy designs entail a plan or framework outlining the aims, objectives and strategies for implementing policies. The goal of each initiative was included in the database.

3.2.2. Data analysis

To select policies that address transformative change, the policy goals of the initiatives were reviewed and filtered by identifying whether the initiatives specifically aim to achieve a green and smart mobility system. For example, initiatives such as 'Geluidsisolatie Schiphol' (noise isolation at Schiphol) do appear in the database but do not contribute to a green or smart mobility system and were therefore omitted from the TIP database. On the other hand, initiatives such as the 'Nationale Agenda Laadinfrastructuur' (national agenda for charging infrastructure) do contribute to this transition and were therefore included in the database. A total of thirty policy initiatives were included which aim to catalyse and accelerate system changes through the promotion and facilitation of innovation. These initiatives were therefore classified as TIPs.

To see whether the Ministry addressed the transition tasks, as described in section 2.2., through the development of TIPs, a document review of the selected policy designs was conducted, which included a coding process of the transition tasks. Prior to the coding process, a codebook was developed that followed the categories of transition tasks and specific transition tasks as presented in Table 2 (see Appendix B for the codebook). To get a comprehensive understanding of the transition tasks before the coding process, the transition tasks were familiarized by closely reading their definitions, descriptions and key characteristics as described by Braams et al. (2021). The policy designs were reviewed and specific attention was paid to sections that discussed transformations, shifts or changes in certain aspects. Through a deductive coding process in NVivo software, the specific transition tasks present in the policy designs were coded.

¹ Dutch Climate Agreement – www.klimaatakkoord.nl/mobiliteit

² <https://open.overheid.nl/documenten/ronl-a5f31aff-596d-4ce9-a83a-b4fbe21925ea/pdf>

To summarize, the analysis of TIP designs provided insights into the adoption of transition tasks in current policy designs. This showed whether the tasks were addressed in policy designs. This was necessary to conduct the following step which analysed whether the transition tasks were subject to evaluation and how they were evaluated.

3.3. Step 2: the evaluation of transition tasks in TIPs

3.3.1. *Data collection*

To see whether and how the transition tasks were evaluated, Step 2 consisted of gathering evaluation reports of the selected TIPs in Step 1. Public search engines, government websites and ministerial digital archives were consulted to obtain evaluation documents. To obtain ministerial records of conducted policy evaluations, the Ministry provided the researcher with a one-hour training in complex software (Content Manager). Six evaluation reports were publicly available and three additional evaluations were found in ministerial digital archives, which led to a total of nine evaluation documents.

3.3.2. *Data analysis*

The evaluation documents were categorised into formative and summative evaluations to see whether the evaluations adopted more learning-based or accountability-based evaluation approaches. The researcher followed the description of formative and summative evaluations as discussed in section 3.3.2. Characteristics of formative and summative evaluations were used as indicators. For example, evaluations aiming to indicate policy 'effectiveness' and 'effectivity' were categorised as summative, and evaluations aiming to learn from 'stakeholder experiences' were categorised as formative. Some evaluations did have characteristics from both formative and summative evaluations. These evaluations, however, emphasized in their evaluation approach that the main focus of the evaluation was learning and were therefore categorized as formative evaluations. Simultaneously, the evaluation documents were deductively coded into transition tasks similar to the coding process described in Step 1.

To summarize, the collected evaluation documents were analysed to identify whether the transition tasks were evaluated in formative and summative evaluation. The objective assessment demonstrated that the transition tasks were evaluated, which was necessary for the sequential step that assesses these observations against expert perspectives in the field of evaluation policy.

3.4. Step 3: designing an evaluation approach and identifying tensions

The third step consisted of interviews to identify whether the transition tasks are recognised by policy evaluation experts as criteria that should be subject to evaluation. Discussing how to evaluate transition tasks led to a suggested approach for TIP evaluation and uncovered tensions in evaluation practices.

3.4.1. *Data collection*

Participants for this study were selected based on their expertise in the field of TIP evaluation. A purposive sampling approach was used to select individuals with in-depth knowledge and extensive expertise in this specific field. To select experts, two criteria were considered. To ensure that participants had a comprehensive understanding of the subject matter and substantial practical knowledge, individuals in senior or leadership positions within their respective organisations were targeted. The selection process prioritized individuals who made significant contributions in their field of expertise through research studies, publications or reports related to TIP evaluation. The selected participants consisted of three groups to capture different perspectives within the field of policy evaluation. The study included academic scholars who conducted extensive research in the field of policy evaluation, practitioners within Dutch government organisations who had experience in designing and reflecting on TIP evaluation, and consultants who had experience in conducting TIP evaluations (see Appendix C for an overview

of the participants). Two participants fell into two participant groups; one academic scholar worked together with government organisations to design TIP evaluations, and one participant working for a government agency published multiple scholarly articles on TIP evaluation. The overlap between groups, although minimal, contributed to a broad range of insights and led to a holistic discussion during the interviews.

The participants were recruited using a similar approach. The pool of potential participants was found through the evaluation documents collected in Step 2, the professional network of the researcher, reviewing literary works, and recommendations from field experts. The individuals were contacted through email or telephone explaining the purpose of the study and the significance of their participation. To ensure diversity in participant groups and enrich the data collection, participants from varying organisations were included in the study. In total, ten experts participated in the study which included four academic researchers, four actors within government organisations and two consultants. The number of participants was selected to prioritize detailed insights and in-depth perspectives from each participant. The seniority within their organisation made their insights highly valuable and each participant provided novel information and unique perspectives that added richness and depth to the collected data.

Interviews were conducted following a semi-structured format (see Appendix D for the interview guide) which balanced between predetermined questions and the flexibility to explore individual perspectives and emerging themes. The interviews discussed three main topics. They provided insights into how field experts experienced the evaluation of transition tasks in current evaluation practices. The discussion on the evaluation of transition tasks gave input on how evaluations should be designed for adequately evaluating these tasks. Combined with participant perspectives on challenges and possibilities in evaluating policies that cope with the complexity and uncertainty of transitions, this research suggests an evaluation approach suitable for evaluating TIPs. During the interviews, there was observed that suggestions for a new evaluation framework differed from dominant evaluation practices. Following this observed difference, the interviews aimed at understanding the barriers to the institutionalization of the discussed framework. This uncovered tensions in how TIPs should be evaluated and how current evaluation practices evaluate TIPs.

All interviews were recorded with the consent of participants and later transcribed for analysis. Each interview lasted about sixty minutes which allowed for an in-depth discussion on participants' experiences, perspectives and expertise. The interviewees were informed about the voluntary nature of their participation and their right to withdraw from of the study at any time without providing any reason. The participants were anonymized in this study.

3.4.2. Data analysis

The interviews were imported into NVivo software for a systematic analysis. The analysis involved a multi-step coding process (see Appendix E for the coding tree) which started with the coding of transcripts. A pre-established codebook was used based on the research objectives and literature review, making the coding process deductive in nature. The codebook included transition tasks, a proposed methods for evaluation, tensions in evaluation practices and necessary conditions for institutionalisation. The transcripts were systematically coded using the pre-established codes. To capture the discussed concepts, each relevant item was assigned to a code. The initial coding processes helped to structure and organise the data accordingly. After the initial coding process, assigned codes were reviewed to identify patterns and commonalities which allowed for the grouping of codes into categories within the established themes prior to the coding process. The data within each category was summarised to capture the main points and insights from the participants. Significant quotes and examples that represented the perspective of participants were extracted.

To summarize, the interviews led to the development of a suggested approach for the evaluation of transition tasks and TIPs in general, and revealed tensions in the field of evaluation. The next step consisted of a case study to demonstrate how the suggested approach would have benefited this particular case. Moreover, the case study identified factors that create favourable conditions for learning as a policy goal.

3.5. Step 4: A case study of the Dutch Mobility as a Service Program

3.5.1. Case selection

A case study on the Dutch Mobility as a Service (MaaS) program was conducted to illustrate the implication of the proposed approach in the previous step. MaaS is seen as a disruptive niche innovation that has the potential to fundamentally change the mobility system (Audouin & Finger, 2018). I&W acknowledged the potential of MaaS and wanted to learn how the concept of MaaS could contribute to policy objectives (Parliamentary Letter, 2018a). Because the program and its evaluation revolved around learning about what MaaS could become, MaaS is a suitable case to demonstrate the added value of the proposed approach that emphasises reflexive learning processes. Moreover, Step 3 identified tensions in the field of evaluation that arise between the accountability and learning function of evaluations. Because MaaS embraced learning as its policy goal, the MaaS case is selected to uncover what conditions allowed learning as an acceptable function of policy and policy evaluation.

3.5.2. Data collection

The data for the case study consisted of five MaaS documents that provided insights into how the program developed and how the program was evaluated. The documents included the official evaluation report³ of the MaaS program, a scholarly article⁴ that conducted a case study on the MaaS program to analyse institutional work from entrepreneurial civil servants, and three parliamentary letters⁵ about MaaS. The official evaluation report of MaaS was already included in the data at Step 2 as an evaluation report of TIPs. The scholarly article was found through scholarly search engines by specifically filtering on the Dutch Mobility as a Service program. By searching publicly available government databases using the keywords 'Mobility as a Service' and 'MaaS', three parliamentary letters were retrieved that reported on the development and progress of MaaS. The documents contributed to the analysis of the proposed method and the examination of tensions related to learning-based approaches. Furthermore, the lead researcher of the scholarly article made an extensive timeline available that supports the findings of that study. The timeline helped for a comprehensive understanding of the MaaS case and its development.

3.5.3. Data analysis

A document review was conducted that allowed for an objective and impartial assessment of events leading to specific outcomes. The document analysis involved a two-step process that shows how the proposed approach could have resolved certain barriers early in the MaaS process and identified favourable conditions that allowed the acceptance of learning as the goal of MaaS.

To analyse how the proposed approach would have led to improved outcomes, moments of contestation regarding the execution of transition tasks were identified. This consisted of revising the codes created in Step 2 from the official evaluation report of MaaS. Codes that showed suggestions for improvements in the MaaS program were included in the analyses as these codes could reveal barriers to the execution of transition tasks. Codes that only showed a positive evaluation of a certain task were omitted. This resulted in five barriers in *give direction* and *support governance*. In addition, the scholarly article was imported into NVivo software. Together

³ <https://open.overheid.nl/documenten/ronl-55a177663787c551e1f2595bd5fcf755da12bc4e/pdf>

⁴ (Braams et al., 2023)

⁵ Kst-31305-260, Kst-31305-271, Kst-31305-294

with the provided timeline, moments of contestation for executing transition tasks were identified and coded to transition tasks accordingly (see Appendix F for coding tree). This resulted in additional information on the four identified barriers and led to one new barrier in *develop internal capabilities and structures*. After the moments of contestation were coded, the proposed evaluation approach from Step 3 was applied retrospectively to the moments of contestation. The assessment analysed how the outcome of each moment that formed a barrier to the execution of each task would have been changed and improved by the proposed approach. This involved the assessment of what actions led to the moment of contestation and how the characteristics of the suggested evaluation approach would have prevented the contestation.

To analyse the favourable conditions that accepted learning as the policy goal of MaaS, the official MaaS evaluation, scholarly article and parliamentary letters were imported into NVivo software for an additional round of coding. The process involved the use of an inductive coding technique, which allowed the researcher to identify statements that related to the learning about MaaS or describe what was necessary to facilitate learning. To ensure that all relevant information was included, the coding process followed an open-ended and exploratory approach without predetermined categories. Similar codes were categorized that described how learning was accepted. The grouping of codes facilitated the identification of common themes and patterns within the data. This resulted in four conditions that favoured learning in the MaaS case.

4. Results

The results are presented as follows. Section 4.1. shows the presence of transition tasks in TIP designs and evaluations, and assesses these observations against expert perspectives on current evaluation practices. Subsequently, based on these findings, Section 4.2. proposes an evaluation approach suitable to the evaluation of TIPs. Section 4.3. discusses the tensions in the field of evaluation as observed by the interview participants.

4.1. Transition tasks in TIPs

4.1.1. Addressed transition tasks in TIP designs

The analysis revealed that TIP designs encompass a variety of specific transition tasks across different transition task categories (Table 4). From the total of 36 designs analysed, *give direction* was the most commonly executed task and was addressed in 25 designs with a total of 120 references across the 25 designs (25/120), followed respectively by *support the new* (22/112), *support governance* (16/38), *develop internal capabilities and structures* (15/38) and *destabilize the unsustainable* (9/21). All transition tasks are addressed in policies aiming for sustainable mobility, however, some transition tasks are more prevalent than others. The analysis showed that the addressed transition tasks relatively vary amongst transformative policy designs, but are all covered by the TIPs.

Table 4: transition tasks addressed in policy design of TIPs

Category of transition tasks	Number of policies including transition task / Total number of references	Specific transition tasks	Number of policies including specific transition task / Total number of references
Give direction	25 / 120	Articulate the direction	21 / 85
		Construct policy strategies in order to direct	5 / 5
		Reconfigure the market	13 / 20
		Direct through enforced regulations	6 / 10
Support governance	16 / 38	Activate actors	4 / 6
		Guiding organizational arrangements	13 / 26
		Goals achieving strategies	4 / 6
Support the new	22 / 112	Engage in entrepreneurial experiments	10 / 23
		Establish market formation	12 / 34
		Price-performance improvements and resource mobilization	17 / 36
		Help new developments develop and diffuse	10 / 19
Destabilize the unsustainable	9 / 21	Control policies and make significant changes in regime rules	9 / 20
		Reduce support for dominant regime technologies	1 / 1
Develop internal capabilities and structures	15 / 38	Rethink own role in a transition	2 / 2
		Development of new competencies	6 / 13
		Monitor and evaluate	9 / 13
		Establish mechanisms for policy coordination	8 / 10

Box 4.1.1: Main findings:

The transition tasks are addressed in transformative policy designs.

The adoption of the transition tasks varies among transformative policy designs.

4.1.2. *Evaluated transition tasks in TIP evaluations*

Looking at how the transition tasks are evaluated in current TIP evaluations showed a difference in the evaluated tasks in formative or summative evaluation approaches. Table 5 presents the number of evaluated transition tasks in the analysed TIP evaluations where the evaluations are categorised into formative and summative evaluations. The analysis shows that *support governance* and *support the new* were the most evaluated tasks in TIP evaluation documents. *Support governance* was evaluated in all formative and most summative evaluations but showed a higher frequency in formative evaluation. *Support the new* was evaluated in all evaluations and showed a higher frequency in summative evaluations. The difference in frequencies of evaluated tasks between the two evaluation approaches can be explained by the characteristics of formative and summative evaluations. Where formative evaluation focused on policy improvement, for example by evaluating the barriers in current governance structures, summative evaluations focused more on policy outcomes, such as the development of new technologies, to measure the effectiveness of the policy. *Give direction* was not evaluated in every evaluation but shows a higher frequency in formative evaluation. *Destabilize the unsustainable* is hardly evaluated and the evaluation of *develop internal capabilities and structures* tends to be more evaluated in formative evaluation but lacks evaluation in most evaluation reports. Overall, higher coverage of the evaluation of transition tasks can be observed in formative evaluation (average of 3.1) in comparison with summative evaluations (average of 1.96). This suggests that formative evaluation approaches are better suited to evaluate transition tasks than summative evaluation.

Box 4.1.2: Main findings

The transition tasks are to some extent evaluated in transformative policy evaluations.

Formative evaluations seem better suitable to evaluate transition tasks.

Table 5: Number of transition tasks addressed in TIP evaluations (see Appendix G for the description of abbreviations)

	Evaluation approach	Give direction	Support governance	Support the new	Destabilize the unsustainable	Develop internal capabilities and structures
MaaS	Formative	3	4	5	0	0
NAL	Formative	6	3	3	0	5
TBOV	Formative	0	8	2	0	0
US	Formative	3	12	2	0	6
DKTI	Summative	0	1	7	0	0
DBL	Summative	2	3	6	1	1
CETSI V	Summative	0	1	9	0	0
SIDB	Summative	1	2	6	0	0
SGV	Summative	1	0	8	0	0

4.1.3. Comparing transition task observations with expert perspectives

In this section, the observations of the evaluation of transition tasks were assessed against perspectives from policy evaluation experts. It analysed how each task was evaluated in policy evaluation and whether the tasks were recognised as evaluation objectives in evaluation practices. Through discussing the evaluation of transition tasks with participants, suggestions emerged on how the tasks can be evaluated in policy programs that foster transformative change.

4.1.3.1. Give direction

The analysis of TIP evaluation documents showed that giving direction is subject to evaluation in both formative and summative evaluation. However, a difference can be observed when looking at how formative and summative evaluations evaluate the tasks. Formative evaluations tend to evaluate the organisational structures to give direction. These conditions include the necessary involvement of actors, current law and regulation and the government's role in further development. These evaluations aim to identify areas where improvements can be made to the conditions for giving direction. In contrast, summative evaluations overall tend to evaluate the actual giving of direction. The evaluations assessed whether a roadmap was developed and if operational barriers were resolved to achieve certain outcomes. These evaluation approaches aim at demonstrating policy achievements and outcomes and can be seen as result-oriented.

Interestingly, when the transition task was discussed with interviewees, there was little recognition of this specific task in current evaluation frameworks. Participants listed challenges regarding its evaluation. For example, evaluating the direction becomes dependent on how one vision a desired future in terms of evaluation indicators and visions are often a societal construct that should not only be determined by governments (Interviewee 4). All the interviewees deemed the tasks important in transitions as “a government has to take a leading role in formulating visions of a desired future and pathways to achieve this” (Interviewee 5). However, many participants described that current evaluation frameworks are inadequate to address or emphasize giving direction as a transition task (Interviewees 3, 4, 5 and 9). One recently developed evaluation framework came up that included giving direction as a component of the transformative capacity of policies (Interviewees 6 and 7). This was addressed by two interviewees who worked together on developing this evaluation framework. Because of the novelty of the framework developed within a specific government department, other participants were not familiar with the evaluation approach.

These findings suggest two gaps in current evaluation practices. First, observations from participants suggest that evaluation frameworks often overlook giving direction as a transition task, while policy evaluation experts recognize *give direction* as an important task. This shows the absence of suitable evaluation frameworks that capture this task which is observed as essential for transitions. Second, *give direction* is evaluated in current evaluations, but existing evaluation frameworks do not specifically emphasise that it should be subject to evaluation. Applied evaluation frameworks do not include giving direction as a criterion that should be evaluated, but the task is, however, evaluated in current evaluations. This misalignment between evaluation frameworks and evaluation practice calls for a more comprehensive evaluation approach. Give direction may be evaluated in current frameworks, but the lack of emphasis on the task in policy assessment may lead to inconsistencies in how the task is evaluated in different policies.

4.1.3.2. Support governance

Comparing the outcomes of formative and summative evaluations regarding support governance reveals a similar difference as described above. Formative evaluation approaches show suggestions for improved governance structures within projects. The MaaS project concluded that more cooperation between parties is necessary to boost the ecosystem, the US that there was insufficient support from external actors, and the TBOV that there is a need for transparency about

governance and certain roles of actors. Although some parts of the evaluations assessed certain governance achievements, most aspects concern suggestions for better governance structures. Summative evaluations evaluated whether the policy contributed to governance structures. The DBL evaluation, for example, concluded that the program led to collaboration and that it mobilized financial actors, and financial schemes, such as the DKTI and the SIDB, evaluated if subsidies led to improved ecosystems.

When the interviewees were asked about evaluating *support governance* in existing evaluation practices, academic participants and participants employed at government organisations highlight that reflexive evaluation approaches are already conducted at a small scale that evaluates how policy processes are organised to assess whether conditions are created that can contribute to a transition (Interviewee 6, 7, 10). These evaluations specifically evaluate governance structures and collaborations with a focus on improving these structures to create a more efficient ecosystem. Through reflexive evaluations one “creates a learning network that develop itself along the way ... so actors know where and how to find each other” (Interviewee 6). It activates actors by encouraging them to participate, facilitates the development of networks and stimulates a collective learning process. Support governance is recognized as an important task by all the interviewees and evaluation approaches exist that are able to capture this task. However, few evaluations are conducted that take this reflexive approach.

These observations suggest that formative evaluations are better suited to address certain barriers to improve governance structures, whereas summative evaluations aim to assess a policy’s effect on structures. Moreover, it reveals that evaluation frameworks are in place to evaluate *support governance*. However, the reflexive evaluation approach was not observed in the analysis of the evaluated TIPs, which only showed formative or summative evaluation approaches. This demonstrates a similar gap as described earlier, where support governance is not specifically emphasized in applied evaluation frameworks, but does seem to be evaluated. This misalignment showed a reoccurring pattern where important aspects of TIPs are not adequately evaluated in existing frameworks.

4.1.3.3. Support the new

The analysis of evaluation documents showed a difference in the focus of formative and summative evaluation. For example, MaaS identified barriers and required conditions for the scale-up of MaaS-apps, the NAL identified institutional barriers that hampered its development and the US concluded that the agenda formed the basis to establish zero-emission zones. The formative evaluations assessed the factors for the successful implementation of programs. On the other hand, summative evaluations tend to evaluate the outcomes of supporting novelties through initiatives. For example, the DBL program increased investors and experimentation in the market, the DKTI scheme boosted investments that otherwise would not have been made and SIDB spin-off effects brought innovations closer to the market.

Again, supporting new innovations that contribute to the transitions was highlighted as important by all interviewees. The same evaluation approach as mentioned by two interviewees in section 4.1.3.1. included promoting novelties within systems as a component of transformative capacity. However, the task was barely discussed by interviewees. Where other tasks gained more immediate attention when they were discussed with the participants, no real challenges were specified by the interviewees for *support the new*. An explanation might be that promoting new innovations is already embedded in traditional frames of innovation policy. Where new challenges arise for the third frame of transformative innovation policy, more traditional evaluation frameworks that originated from the first and second frame of innovation policy are focused on growth and therefore included promoting new innovations as a growth indicator. Therefore,

support the new was less touched upon as it could be seen as a less emerging challenge for evaluating TIPs.

These observations show that current evaluation approaches are capable of evaluating the promotion of new innovations and that assessing this task is not observed as a direct challenge. This could indicate that current evaluation frameworks are based on more traditional frames of innovation policy, which already emphasise promoting new innovations. Therefore, support the new is not seen as an emerging challenge for new evaluation frameworks. However, this task is an essential element of TIPs and new assessment frameworks should therefore not overlook promoting novelties as a critical function for transitions.

4.1.3.4. Destabilize the unsustainable

Breaking down the regime was hardly addressed by any of the evaluations. Only the DBL evaluation shows that existing practices are pressured by assigning research that investigates admixing obligations for sustainable aviation fuels. This is in line with the analysis of addressed transition tasks in policy designs where destabilizing the unsustainable is the least addressed transition task.

Interviewees addressed that pressuring the status quo is something that is often neglected in policy making and evaluation because of the resistance one encounters in destabilizing dominant practices (Interviewees 4, 5, 7 and 8) . “it asks for political courage. And in that sense, politicians have become a bit frightened of citizens.” (Interviewee 8). Accordingly, participants did not mention existing evaluation frameworks or approaches that addressed or evaluated this task. However, to create room for new innovations and change dominant systems, all interviewees acknowledged that pressuring regime structures is an essential process for transitions. Suggestions were made on how this task could be more adopted in policy evaluations: “evaluation should not only include what is accomplished but should also evaluate what remains untouched.” (Interviewee 7). Evaluations should “evaluate whether we have already phased things out in this transition” (Interviewee 2). Current evaluations tend to capture what changes current or new TIPs evoke in existing systems. However, parts of these systems remain untouched by TIPs while these parts are existing forces that keep the sociotechnical system in place. Evaluating what remains untouched by policies can map opposing forces that hamper new developments and make TIPs less effective. In doing so, TIPs can be targeted at these opposing forces and accelerate sociotechnical change.

These findings suggest that there are no evaluation frameworks in place to assess the destabilization of unsustainable practices, nor do current evaluations capture this task. Because this task is an essential process for transition, new comprehensive evaluation approaches should be developed that emphasize destabilizing current practices.

4.1.3.5. Develop internal capabilities and structures

The formative evaluation of the US reflected on its internal government processes by highlighting that it lacks a unified program team within the Ministry and that intensified collaborations within the Ministry led to better coordination. The NAL showed that a lack of trust led to parallel operating activities and the inadequacy of coordination complicated decision-making. These outcomes all addressed specific barriers that hampered the development of the programs to improve policymaking. The summative evaluation of the DBL points out that ministerial resources were made available to fulfil program ambitions and so evaluated what was done to improve the program, rather than address areas for improvement.

The tasks came up in multiple interviews and were addressed as an important aspect of transitions. “Before that [system change] can happen, institutions responsible for policy

development have to transform themselves” (Interviewee 7). Evaluations in this context have to evaluate TIPs in a way that suits existing institutions but on the other hand, transform these institutions accordingly. Interviewees with an academic background highlight that an evaluation of the transformative capacity of policies should evaluate how the organisation of internal structures allows governments to fulfil other tasks (Interviewees 6, 7 and 9). It should evaluate whether the conditions are created that allow governments to give direction, support governance, support the new and destabilize the unsustainable. However, there was little recognition of this task in current evaluation practices. The two consultants highlighted that they assessed governance structures for policy evaluation, but did not specifically evaluate how internal government processes were organised (Interviewees 1 and 2).

Current evaluations do seem to evaluate internal capabilities and structures in some programs, but it is not consistently evaluated across the various TIPs. The inconsistency and limited focus on internal capabilities and structures can hindered the execution of other transition tasks. New evaluation frameworks should be designed as an approach that includes and explicitly evaluates the development of internal capabilities and structures to enable external tasks.

The following can be concluded from the analysed of transition tasks in TIP evaluation reports and interviews. Overall, interviewees mention that the complexity of TIPs makes it challenging to assess the transition tasks in terms of their achievements, for example, evaluating to what extent direction is given as an effect of the policy. This relates to the longer time scales on which TIPs operate making it difficult to evaluate outputs. Evaluating TIPs should go beyond assessing outputs and outcomes that are currently evaluated by summative approaches. Instead, TIP evaluation should prioritize learning and adaptivity as a function of evaluation by addressing areas for improvement. It should provide insights into what works and what does not work in the policy process. This asks for “adjusted questions subject to evaluation. ... the emphasis should be on the contributions of policy to a transition” (Interviewee 3). Through a process of learning and policy steering as a result of evaluation, factors can be assessed that hamper or stimulate the execution of certain transition tasks. TIP evaluations should evaluate whether the right conditions are created (or not) to enable the execution of transitions tasks and identify areas for improvement to better execute these tasks.

This goes, however, beyond solely formative evaluation approaches. While the analysis of TIP evaluation reports show that formative evaluations emphasize learning and policy improvement, they do not necessarily facilitate critical reflection on unintended effects, power dynamics and underlying assumptions. Critical reflection encourages stakeholders to challenge existing norms and dominant structures within the system context. Evaluating how conditions can be improved to successfully execute transition tasks asks for a deeper examination of resistance and systemic barriers that only formative evaluations are unable to capture. By evaluating the organisation around tasks early in the policy process, policies can be adjusted and “navigated in the insecurity of transitions” (Interviewee 9), which enhances the transformative capacity of TIPs. The following section describes how an evaluation approach should be designed that allows for the evaluation of these conditions.

Box 4.1.3: Main findings

The contribution of transformative policies should be evaluated by assessing whether the right conditions are created to successfully execute transition tasks.

The evaluation of conditions asks for critical reflection, going beyond solely formative evaluations

4.2. A suitable approach for TIP evaluation

The following section describes how an evaluation approach can be designed to evaluate TIPs. It combines the suggestions on how to evaluate transition tasks with insights from interviewees about evaluation challenges in transition to propose an evaluation approach suitable to the evaluation of TIPs.

4.2.1. A reflexive evaluation approach

All the participants highlighted the need for new evaluation approaches that are able to capture the complexities of TIPs. They stressed the urgency for a new type of evaluation approach that is not only focused on accountability and justification but should emphasize learning and adaptivity through reflexive processes and stakeholder inclusion. The majority of the interviewees highlighted that more formative evaluations are necessary to be adaptable and flexible in complex TIP processes (Interviewees 3,4, 5, 6, 7, 9 and 10). Participants who have expertise in specifically learning-oriented evaluations highlight that a reflexive evaluation approach that revolves around second-order learning to improve the policy process suits the learning-based function of evaluation in uncertain transitions (Interviewees 3, 6, 7 and 10). Second-order learning allows participating stakeholders to question underlying assumptions and existing practices, which is relevant for the evaluation of certain conditions that enable transition tasks. By evaluating the broader context and systemic factors that influence certain outcomes, a reflexive evaluation approach takes a holistic perspective that allows for evaluating whether the right conditions are in place to enable the execution of transition tasks. It becomes important to note that the earlier reflexivity is introduced in the policy process, the greater the impact on the policy and its implementation. Starting with a reflexive evaluation approach from the start of the policy process enables policymakers to identify and address barriers in the execution of transition tasks at an early stage. It minimalizes ineffectiveness by adjusting and adapting the policy process real time.

However, interviewees with an academic background argue that a new paradigm of policy evaluation does not mean that the accountability and justification function of evaluation should be abandoned as it remains an important function of evaluation (Interviewees 5 and 7). It fosters trust between the government and the public as citizens can hold policymakers accountable for their actions. Moreover, summative policy evaluations might be able to capture certain aspects of TIPs. A balance must be found regarding more formative and summative evaluations. Interviewees stress that this asks for a shift towards more learning bases evaluations and highlight that learning as a function of evaluation should gain more legitimacy in the evaluation practices.

4.2.2. An inclusive and participatory process

Regarding the evaluation type for TIPs evaluations, interviewees mentioned that in reflexive evaluations a participatory approach is the heart of the process. These evaluation processes benefit from engaging different stakeholder views in policy assessment. TIPs are aimed at promoting system change and often require new ways of thinking, new forms of collaboration, and new approaches to governance. These types of policies involve complex and uncertain systems, where the impacts of the policy are not always easy to predict. A constructivist evaluation approach is well-suited to navigate these complexities by acknowledging and embracing the diversity and complexity of the social context in which the policy operates. "It is, therefore, necessary to map all possible effects, intended and unintended, which requires including lots of stakeholders across different domains. TIP evaluation becomes a very holistic story" (Interviewee 3). A holistic perspective that includes stakeholders in the evaluation approach allows for the understanding of underlying processes, dynamics and assumptions that shape the transition progress. As such, an inclusive and participatory evaluation approach becomes desirable for TIP evaluations. Furthermore, to allow second-order learning, stakeholders must critically reflect on

their assumptions, values, and beliefs about the policy or program being evaluated. The engagement of stakeholders in this type of reflection can identify areas for improvement and modify the policy or program in ways that are more responsive to the needs and perspectives of those who are affected by it. Therefore, an inclusive process benefits the evaluation of conditions for transition tasks. In this way, second-order policy learning can lead to more effective policies and programs.

4.2.3. *Developing a flexible theory of change*

In addition, interviewees mentioned another important aspect regarding TIP evaluation: developing a flexible Theory of Change (ToC) (Interviewees 3, 5, 6 and 8). A ToC provides the logic of how the inputs invested in a policy are expected to lead to a set of outputs and relevant outcomes. A ToC is usually developed ex-ante by identifying the main changes one is trying to achieve. Through a process of back-casting, policy stakeholders develop a logic of what processes will lead to the desired changes. It provides policymakers with a systematic and strategic framework that articulates policy objectives, map out pathways and identifies potential risks and challenges. Through the development of a ToC, TIPs can be grounded in evidence-based reasoning that stimulates coherence in policy design and its implementation. Interviewees specifically stress the importance of a *flexible* ToC, meaning that the developed ToC should be revised and adjusted alongside the policy process. TIPs often cope with complex and dynamic systems. A flexible ToC acknowledges the complexity and uncertainty of changing systems and enables policymakers to adjust the developed theory as new information emerges. “With our present knowledge, we cannot predict how a certain policy will unpack in a transition. Therefore, it becomes necessary to develop a theory of change, and continuously revise if the theory is still accurate.” (Interviewee 8). It allows for agility in responding to emerging challenges, opportunities and feedback mechanisms that help to improve policies in fostering systemic change.

The development and adjustment of a flexible ToC should be an inclusive process that includes all relevant stakeholders, such as industry experts, community representatives, researchers and policymakers. By incorporating a variety of perspectives, expertise and knowledge, a constructivist approach enhances the legitimacy of the developed ToC by ensuring that the theory accurately represents the complexities and needs of the targeted system. The flexible ToC will foster learning and reflexivity among stakeholders and helps to assess if a policy is contributing to moving towards its objectives. “The critical reflection among stakeholders pushes second-order learning.” (Interviewee 3).

The primary goal of the proposed evaluation approach is to evaluate whether conditions are created that enable the successful execution of transition tasks. This involves assessing whether policy interventions and implemented activities align with the execution of transition tasks as transformative outcomes. Through the development of a flexible ToC, one can revise and adapt the ToC based on new insights and feedback from evaluation processes. It enables evaluators to assess the alignment between expected implementation and actual context, to examine if conditions are created to execute transition tasks.

Box 4.2: Main findings

An evaluation approach that evaluates whether conditions are created to execute transition tasks should be:

- Reflexive – revolve around second-order learning
- Participatory – includes stakeholders
- Based on a flexible ToC – revise and adjust theoretical assumptions

4.3. Tension in evaluation practices

Despite the consensus for more learning-based approaches to evaluate TIPs, evaluation practices are too focused on accountability and justification, often in the early stages of the policy process where “There is much attention for the part that cannot yet be assessed” (Interviewee 5). Summative evaluations were conducted that aim to assess the effectiveness and efficiency of TIPs where little to no effects can yet be evaluated. Formative or reflexive evaluations are more suitable to evaluate policy processes and examine whether the conditions are created that enable the execution of transition tasks. In the field of policy evaluation, a mismatch is observed between what evaluation approaches are needed for TIPs and how evaluations are currently executed.

An interviewee explains that the role of the civil service towards the Parliament regarding evaluations currently rests on the justification of policy effectivity and efficiency: “This means that evaluations and the current evaluation regime within ministerial departments are based on these [justification] criteria” (Interviewee 7). It is, however, difficult to shift towards formative evaluations that allow for learning and adjusting. “The justification function is crucial in our representative democracy ... when a reflexive evaluation creates feedback loops and adjust goals, how do you maintain the justification policies?” (Interviewee 7). As policy goals and implementation are revised, it becomes challenging to justify these policies in the end, since it becomes difficult to determine which goals should be evaluated. Moreover, learning as a function of policy evaluation involves experimentation and taking risks, which may result in unexpected outcomes or ‘failures’. While failures are an inherent part of the learning process, in the context of accountability, policymakers are responsible for these negative outcomes. The fear of being held accountable for the failure of TIP interventions can discourage policy actors to embrace learning as a function of evaluation.

This puts pressure on the role of civil servants to ask for formative or reflexive policy evaluations. These reflexive evaluations require civil servants to bring stakeholders together and collaborate with policy actors to learn instead of evaluate effectiveness. This “asks for agency from civil servants that do not always fit their job descriptions, which creates tensions” (Interviewee 7). Institutional government structures ask for effectiveness and accountability from civil servants making the institutional context civil servants operate in not adequate for policy learning. “In a hierarchical bureaucratic system, what a civil servant does goes through the minister to the Parliament. So the internal accountability structure is one to take into account.” (Interviewee 6). The internal structures are not designed to execute and evaluate TIPs reflexively. “that space to stop for a moment and reflect on the execution of policies is not felt by the policy officers who are involved in the implementation of the policy file.” (Interviewee 6). Where transitions ask for reflexive evaluations of TIPs to navigate in uncertain transitions, government organisations do not necessarily provide room to ask for and conduct these learning-based evaluations. Governments must transform themselves to legitimize civil servants’ actions in evaluating TIPs accordingly.

Box 4.3.: Main findings

The shift towards learning is difficult because the accountability function of evaluation remains important in our representative democracy.
Being transformative within government organisations is often contrary to what is expected from civil servants.

5. Case study

This section set out to demonstrate the implications of the proposed approach and identify favourable conditions for learning. The MaaS program revolved around learning what MaaS could become (Parliamentary Letter, 2018a). Despite this similar goal of MaaS and the proposed approach, the MaaS program was officially evaluated after five years and concluded that “resolving most barriers and creating conditions for MaaS still have to take place” (TwynstraGudde, 2022, p. 30). This case study demonstrates whether these barriers would have been resolved following a reflexive evaluation approach. Moreover, as MaaS revolved around learning, the case study is used to identify the conditions that allowed the program to accept learning as its goal.

5.1. The implications of the proposed approach

The results of the case study are presented as follows. For each transition task, moments that led to contestation are presented that formed a barrier for executing the transition task. The context that hampered the execution is described whereafter is discussed how the proposed approach could have changed and improved the conditions that hamper the execution of the tasks. Barriers to the execution of transition tasks were found in the transition tasks *give direction*, *support governance* and *develop internal capabilities and structures*.

5.1.1. Give direction

Three points were identified that hampered the execution of giving direction as a transition task. First, MaaS requires the creation of stable policy frameworks regarding guidance and market formation. A barrier emerged when the national railway company was unwilling to extend their 40% discount, deployed to manage rush hours, to MaaS providers who needed to resell railway tickets (Braams et al., 2023). It has become clear that MaaS is primarily a market regulation issue, and that this requires policy and regulations as well as additional forms of market supervision. “This was perhaps already known at the front of the program, but the urgency of it was not felt equally strongly by everyone at the time.” (TwynstraGudde, 2022, p. 29). The disparity of urgency about the market regulation issues encountered during the program eventually hampered the development of the program and inhibited the government from giving direction. A reflexive evaluation approach that revolves around second-order learning and includes all stakeholders early in the policy process would create opportunities for stakeholders to address this disparity of urgency early in the policy process. Through continuous reflection and ongoing dialogue, stakeholders could collectively understand the importance of market regulations. Specifically, the reselling of railway tickets and the unwillingness of the national railway company to do so is a barrier that could be addressed earlier in the policy process. Policy actors would have engaged in critical thinking and found that MaaS providers needed to resell railway tickets as a crucial aspect of the MaaS system. Through stakeholder engagement and actively seeking input and diverse perspectives, the reflexive evaluation would have created an opportunity for stakeholders to raise concerns or challenges related to ticket reselling. These early insights would have triggered a further investigation to understand the implications of the barrier and its effect on the success of MaaS.

Second, policy reforms were hindered because the business case of the national railway company had to be adjusted. Current concession constellations had to change to create a MaaS market. However, the Ministry of Finance already built in expected returns from its current way of operating (Braams et al., 2023). The Ministry was partly unwilling to reconfigure the market and create instability in the mobility system, due to the novelty of MaaS with uncertain outcomes. The resistance to change and preserving the status quo prevented the government from effectively directing and facilitating policy reforms. Although the suggested approach would not have changed the expected returns from the Ministry of Finance, a reflexive approach would have

identified the need for an adjusted way of current operations. Stakeholder engagement could have facilitated the development of mitigation strategies to collectively come up with alternative solutions.

Third, the government should take a leading role in bringing MaaS further. This starts with the political will to take MaaS further and give MaaS a place in the mobility system. This requires a clear view of what MaaS is: an extension of public transport or a private activity (TwynstraGudde, 2022). The absence of this vision led to several conflicts with the national railway company which had more a conservative view on MaaS (Braams et al., 2023). The conflicting perspectives and variety of interests among stakeholders hindered the government's ability to resolve tensions and give cohesive direction. A reflexive approach would have emphasized the need for a clear vision for MaaS, its purpose and objectives, and the role of the government in achieving them. For example, the reflexive approach could have recommended that the program should first focus on creating a shared understanding of MaaS before moving forward with its implementation. This would have led to a more coordinated approach to developing a healthy MaaS market. Stakeholders would have had a clear understanding of MaaS' goals, strategies and responsibilities for achieving desired outcomes. This could have ensured that the government had a clear understanding of the needs of MaaS and what could have been done to contribute to its success.

5.1.2. Support governance

Two points were identified that hampered the support of governance throughout the MaaS program. First, despite being one of the goals of MaaS, the program was unable to create a learning environment due to privacy issues and business confidentiality (Braams et al., 2023; TwynstraGudde, 2022). MaaS providers saw the Ministry as a neutral party to collect data and create a learning environment. However, legal issues arose for the Ministry as an actor responsible for data processing (Braams et al., 2023). A second opinion confirmed the difficulty with the legal basis for the Ministry as data manager but suggested exploring other forms for creating a learning environment (Braams et al., 2023; TwynstraGudde, 2022). The legal problems around data protection regulations were encountered after the first pilots went live. However, questions were already addressed at the start of the MaaS program where the legal department "had questions if the Ministry could take responsibility for processing all the data." (Braams et al., 2023, p. 9). It suggests that this barrier was not accounted for earlier in the policy process. Following a reflexive approach, this issue would not have only been addressed in the processes but would have thoroughly discussed this problem with stakeholders before moving on to the pilot phase as data sharing was necessary to achieve to learning goal of the program. As suggested by the second opinion, the program would have searched for another form for creating a learning environment earlier in the process. Whether the reflexive evaluation would have resolved the problem or not, discussing the problem with all stakeholders would have managed expectations about the learning environment.

Second, public and private transportation providers were not willing to provide their services through MaaS apps. Public transportation providers are not convinced that the number of travellers will increase with MaaS and shared mobility providers do not believe that MaaS apps will attract new customers (TwynstraGudde, 2022). A reflexive evaluation could uncover ways in which the MaaS program could be redesigned to meet the needs and expectations of public and private transportation providers. Through a collaborative approach, their necessary changes would have been discussed that would make MaaS more attractive for the transportation providers. This approach would help to overcome barriers by creating the right conditions for collaboration and innovation that lead to a more successful MaaS ecosystem.

5.1.3. *Developing internal capabilities and structures*

Internally, the MaaS program struggled to get the public transport and railway directorate on board the program. The directorate had other priorities due to their daily business and to them, MaaS sounded like wishful thinking because of its comprehensiveness (Braams et al., 2023). The different priorities within the Ministry led other directorates to question the priority of the MaaS program which resulted in a lack of broad internal support (Braams et al., 2023). The suggested approach that revolves around second-order learning would encourage stakeholders, in this case specifically the public transportation and railway directorate, to critically reflect on themselves. This would include challenging the underlying presumptions, biases and mental constructs that led them to perceive MaaS as wishful thinking. Stakeholders would learn about their perceptions and limitations in their thinking. This increased open-mindedness could facilitate a willingness to consider alternative perspectives, which enables the public transport and railway directorate to see the transformative potential of MaaS and its possibility to change the mobility system.

Box 5.1.: Main findings

A reflexive approach that is implemented early in the policy process would have addressed various barriers. The approach would identify and mitigate barriers early in the process, stimulate collaboration and alter solutions, and foster a common understanding of the transformative potential of MaaS by stakeholders. This would improve conditions for successfully executing transition tasks.

5.2. *Favourable conditions for learning*

The MaaS case embraced learning about the possibilities of MaaS as its main goal: “it emphasis is about the joint acquisition of learning experiences with MaaS on a large scale by MaaS providers, carriers and governments.” (Parliamentary Letter, 2018a). Four potential factors were found that led to the acceptance of learning about Maas, rather than solely focusing on its outcomes and impact.

First, MaaS had the potential to fundamentally change the mobility system which led to an optimistic feeling about the program (Braams et al., 2023). This feeling was combined with a form of urgency to learn about what MaaS could become before big investors could take over the market (Braams et al., 2023). The combination of the novelty and potential of MaaS combined with the urgency and eagerness to learn more about MaaS’ potential facilitated an environment that caused the government parties and the private sector to be open to learning through experimentation. MaaS demonstrated commitment to learning through its pilot projects: “To maximize the learning effects, the pilots have different goals and learning objectives.” (Parliamentary Letter, 2018a). The program specified these objectives as gaining insights into the societal impact of MaaS, the creation of a business case and market development, consequences for policies and possibilities in steering, and the operation of a MaaS ecosystem (Parliamentary Letter, 2019). Through evaluating these objectives, learned lessons can justify the program. The urgency for learning from the pilots was highlighted by opting for “a few scalable national pilots to compare outcomes and prevent a rudimentary monopoly installation to learn as fast as possible” (Braams et al., 2023, p. 8). In this case, it was the combination of the potential of MaaS combined with a feeling of urgency to develop MaaS that caused learning to be the prior goal of the program.

Second, the MaaS program took a collaborative governance approach. MaaS emphasized “collective learning, collaboration and data-sharing” (Parliamentary Letter, 2018b, p. 1), which was supported by many actors. The learning function of MaaS was widely embraced by stakeholders: “it emphasis is about the joint acquisition of learning experiences with MaaS on a large scale by MaaS providers, carriers and governments.” (Parliamentary Letter, 2018a, p. 1). The

Minister reported to the parliament that “we look forward to, together with regional governments and market parties, learn and experiment with MaaS to see if its potential can become reality” (Parliamentary Letter, 2018a, p. 4). Stakeholders were actively engaging in decision-making about the future of the program and felt responsible, creating some sort of ownership in the program. The participatory approach that engaged stakeholders in collective learning emphasized that learning was the main objective of MaaS and highlighted the shared view of the potential of the program.

Third, in the parliamentary letters, the uncertainty of the future of MaaS was acknowledged. The great potential of MaaS combined with the recognition of uncertainty favoured a learning-oriented approach. The Minister highlighted in this letter to parliament that to find out if the potential of MaaS can be realized, there must be experimentation through pilots: “If the conceptual promises are fulfilled in practice, this will offer many opportunities for a more data-driven mobility policy. But these are pilots and things can go wrong in pilots.” (Parliamentary Letter, 2018a). A consecutive letter even emphasized the necessity of the ‘failure’ of pilots: “At this stage, however, we would like to emphasize that these pilots should fail in order to learn from them.” (Parliamentary Letter, 2018b). The Minister specifically emphasized that ‘failure’ is necessary to learn. By recognising uncertainty and acknowledging failure, learning as an outcome of failure can be justified.

Last, political will and support were created. Early from the start of MaaS, the Minister and Ministry were convinced that MaaS should revolve around learning (Braams et al., 2023). This is a crucial factor as the commitment to understanding, improving and willingness to change the mobility system, rather than holding individuals accountable for shortcomings in the program. The Minister and Ministry committed to learning by reporting lessons from pilot evaluation to the parliament (Parliamentary Letter, 2018a). Moreover, the Ministry was designated to be responsible for the learning environment created (Braams et al., 2023). By assigning this responsibility to the Ministry, it indicates a level of commitment to creating a learning environment for MaaS. The intentions to report MaaS lessons, acceptance of responsibility and commitment to learning indicate that political support was created for the continuous improvement of MaaS through a learning-based approach.

Box 5.2: Main findings

Four conditions led to the acceptance of learning about MaaS:

- Seizing the potential
- Collaborative governance approach
- Recognition of uncertainty
- Political will and support

6. Towards a learning-based evaluation framework

The fundamental incompatibility between evaluation practices and the evaluation approaches transitions ask for led to several barriers to adopting and implementing learning-based evaluations. It remains difficult to emphasize learning from a historical perspective because evaluation practices are in place to hold governments accountable for their actions. Within this institutional context, it becomes difficult for civil servants to act transformative. Reflexive evaluations need to gain legitimacy in evaluation practices where learning-based evaluations should be seen as a valuable tool for organizational learning that improves policy development. As mentioned by several interviewees, the current culture of evaluation is often focused on demonstrating accountability and justifying actions and decisions (Interviewees 4, 5, 6, 7 and 9), rather than promoting learning and improvement as “reflexive evaluations are seen as second best” (Interviewee 3). Interviewees highlighted the importance of an institutional culture of learning which asks for changes in structures, processes and mindsets. Combining the findings from interviews with insights obtained from the case study led to several recommendations for enhancing the legitimacy of learning-based evaluations.

Because accountability is a vital function of evaluation, these recommendations promote learning while also ensuring accountability. On a national level, other financial structures can be adjusted to allow for policy learning as a function of evaluation. Existing structures develop indicators for evaluation that are focused at assessing policy outputs and do not justify learning an indicator for policy success. Financial structures can be adjusted to be tailored to “reduce the tensions between learning and justifying on a national scale, for example, by providing a lump sum way of financing” (Interviewee 7). Such financial structures would provide a predetermined budget that allows policymakers and evaluators to focus on learning without being held accountable for every action. To remain the accountability function in new financial structures, we should “Ask parties to develop their own indicators for justification and accountability that they find important ... These indicators can be substantiated with a narrative.” (Interviewee 7). This promotes ownership and allows policymakers and evaluators to align evaluations with their transformative objectives. The narrative provides contextual information and explanation for the developed indicators, which can be reviewed by parties ensuring government accountability. A balance must be found in what learning objectives could be justified through a narrative and when a learning-based narrative is considered ‘good enough’ to be justified.

At a ministry or directorate level, evaluation practitioners and policymakers should feel that reflexive approaches are a legitimate option for TIPs and feel comfortable asking for them. From a top-down perspective, this requires support from organisational leaders. Political will and support allowed MaaS to revolve around collective learning and a formative approach to its evaluation. This top-down support also led to other evaluations that more specifically emphasized reflexivity: “it was the Secretary-General within the ministry that said: “we are going to do things differently” and put standard procedures up for discussion, which eventually led to multiple reflexive pilots.” (Interviewee 10). Learning from evaluations to assess policy contributions in transitions requires organisational leaders to ask and push learning-based approaches.

From a bottom-up perspective, entrepreneurial actors must ask for reflexive evaluations from their superiors. Legitimacy for learning-based evaluation approaches requires the education of civil servants and stakeholders on the principles and methods of reflexive evaluation to create awareness and understanding of the approach. In the MaaS case, the acceptance of the goal to prioritize learning shows that all parties saw the benefits of a learning-based approach. For reflexive evaluations, education about the additional benefits becomes even more crucial as it encourages critical self-reflection which asks for stakeholder commitment. The outcomes of the interviews show that few evaluation frameworks exist that include reflexivity and that these

frameworks are merely adopted in evaluation practices. Only a few pilots and reflexive evaluations recently emerged across different Ministries (Interviewees 6, 7, and 10). To increase the adoption and implementation of reflexive evaluations to assess if transformative policies contribute to system change, policymakers, evaluators, stakeholders and management should be aware of the additional value of the evaluations.

Moving towards a culture of learning, however, asks for starting with a learning-based evaluation of all these aspects: “we actually have to get that ball rolling on all different fronts” (Interviewee 10). To push learning-oriented evaluations, it asks for agency from both leaders who should provide room for learning, and civil servants who should ask and create room for reflexive evaluations. Through educating both civil servants and managers, legitimacy for learning can be developed and created internally. To create legitimacy externally, it becomes increasingly important to promote openness and transparency about learning-oriented evaluations. Policymakers should emphasize the lessons that are learned from policies and more important, be transparent about policies that did not work out as planned. Policymakers should stress the importance of the lessons yielded from ‘failed’ policies to learn and improve future policy development.

Box 6: Main findings
Financial structures should be adjusted to allow for learning.
Learning asks for top-down support from managers.
Educating relevant actors about reflexive evaluation shows the additional value of these approaches.

7. Conclusions

In synthesizing the transition literature that attributes transition tasks to governments with policy evaluation literature, this study aimed to develop a suitable approach to the evaluation of Transformative Innovation Policies (TIPs) and to derive the necessary conditions to institutionalize this approach within the policy evaluation context. At present, evaluation frameworks do not specifically emphasize the evaluation of transition tasks, although these tasks are to some extent evaluated in evaluation documents, this is mostly implicitly. As this leads to inconsistency in the evaluation of transition tasks, an evaluation approach is proposed that is suitable for the evaluation of transition tasks in transformative innovation policies. The approach evaluates whether the rights conditions are created to successfully execute the transition tasks by governments through a reflexive process that revolves around second-order learning and stakeholder inclusion and is based on a flexible Theory of Change. This study finds that current barriers to successfully executing transition tasks can be resolved when the reflexive approach is implemented early in the policy process.

This study set out to identify tensions that hamper the institutionalization of learning-oriented evaluation approaches. Academics already pointed out that it is likely that tensions arise in the field of evaluation with the rise of TIP as the third frame of innovation policy. While academic scholars argued that policies aiming for transformative change cope with the uncertainty of transition and therefore ask for more learning-based evaluation approaches, policymakers are historically bound to approaches that fulfil the accountability function of evaluation. The results show that internal government structures are in place that hold civil servants accountable for their actions rather than providing room for learning. This asks for agency from civil servants that go beyond the traditional normative frames that provide legitimacy to their actions. This finding is in line with literature that highlights that current public administration traditions providing legitimacy to the role of the government do not fit the execution of transition tasks by civil servants. The proposed alternative evaluation approach is a first step in developing a learning-oriented evaluation framework that includes the evaluation of transition tasks. Applying this approach to the case of the Dutch Mobility as a Service program, conditions were identified that allowed the justification of learning. To narrow the gap between accountability and learning, this study also finds that there is a need for other financial structures that allow for learning rather than accountability, top-down support from organisational leaders and the education of actors, to create legitimacy for reflexive evaluation practices.

8. Discussion

Through combining two strands of literature with empirical observations and in-depth insights from experts in evaluating transformative policies, this study proposed an evaluation approach and illustrated first steps in moving towards learning-oriented evaluation frameworks. This section discusses the contributions to literature and addresses the limitations of the study. These limitations are translated into suggestions for further research to build a comprehensive evaluation framework for transformative innovation policies.

8.1. Theoretical contributions

Transformative outcomes as proposed by Schot et al. (2019) and Ghosh et al. (2021) offer direction for the transformative change policies need to trigger. These outcomes follow three processes central in the Multi-Level-Perspective: building and nurturing niches, expanding and mainstreaming niches, and unlocking and opening up regimes (Ghosh et al., 2021; Schot et al., 2019). Lazarevic et al. (2022) argue these processes fall short in phasing out regime structures which is currently not captured by the transformative outcome framework. They suggest the expansion of the framework by including the process of repercussions of regime destabilization, policy coordination, and tilting the landscape. Although these processes are rather similar to the categories of transition tasks as proposed by Braams et al. (2021), this study proposes to explicitly incorporate the transition tasks attributed to governments in the transformative outcome framework. The contributions to literature are twofold. First, this study argues that transition tasks should be adopted as policy objectives. When governments become responsible for the execution of transition tasks to address societal challenges (Braams et al., 2021) and governments should intervene through innovation policy when society fails to become transformative themselves (Weber & Rohracher, 2012), empowering civil servants to execute transition tasks should become a policy objective in itself. By acknowledging transition tasks as policy objectives and therefore transformative outcomes, this study emphasizes the active role of the government in facilitating transformative change and sustainable transformations. It aligns with the broader understanding of transformative innovation policy, which implies that innovation should be aimed at grand societal challenges rather than for the mere sake of innovation or economic growth (Diercks et al., 2019; Grillitsch et al., 2019) and acknowledges that governments need to adopt a societal policy agenda (Diercks et al., 2019). Including transition tasks as transformative outcomes highlights the importance of government involvement and stresses their responsibility. It emphasizes the need for government action and coordination to achieve policy objectives.

Second, this research explores the tensions and challenges that the shift towards the third frame of innovation policy creates for the evaluation of transformative innovation policies. The implementation of these transition-oriented, long-term policies asks for more learning-oriented approaches to evaluation (Janssen et al., 2022; Molas-Gallart et al., 2021). The interviews show the awareness of a need for a shift in evaluation practices that suit TIP challenges. There is a shared recognition of the importance of broader goals and processes related to transformative change, such as engagement with diverse stakeholders and fostering learning and experimentation. This suggests that there is a growing recognition in the field of TIP evaluation of the need to adopt more holistic approaches to evaluation that capture the complexity and dynamics of transformative change. Rohracher et al. (2023) find that challenges in the evaluation of TIPs are caused by an incomplete shift towards the third frame of transformative innovation policy. There is a coexistence of traditional elements and transformative innovation approaches and their respective summative and formative types of evaluation (Rohracher et al., 2023). Following Diercks et al. (2019) who argue that the three frames of innovation policy should be layered on top and not replace each other, the integration of traditional evaluation approaches with learning-oriented policy evaluations would fit the third frame of transformative innovation policy (Arnold, 2004; Janssen et al., 2022). Although this study proposed a reflexive approach to the evaluation of transformative innovation policies, it can be argued that summative evaluations should not be abandoned as they hold the crucial accountability function within our society. These

summative evaluations can still be applied to individual instruments with the purpose of accountability (Magro & Wilson, 2019). Nevertheless, the evaluation of transformative innovation policies should move from being primarily about accountability to being a strategic tool for policy learning and improvement. This study argues that the shift towards the third frame of innovation policy led to the rise and legitimisation of transformative innovation policies oriented towards societal challenges, but dominant evaluation practices are still rooted in traditional frames of innovation policy that favour accountability rather than learning.

8.2. Limitations and suggestions for further research

This research showed conditions that allowed for learning in an accountability-dominated culture. Although the proposed approach is derived from discussions with a variety of experts in the field of policy evaluation, the case study is conducted in a specific policy context within the Ministry of Infrastructure and Water Management. Therefore, the identified favourable conditions that allowed for learning might be not directly applicable to other contexts. Extensive research can be done in identifying the necessary conditions for learning, for example, in the form of other case studies. Further research can conduct a comparative analysis between government organisations that have successfully implemented learning-oriented evaluation frameworks and organisations that have not. The comparison of different cases can provide insights into organisational characteristics and contextual factors that enable or inhibit governments from institutionalizing learning-oriented evaluation frameworks. Research can build a comprehensive framework that analyses the shift towards learning-oriented evaluation approaches that follow the rise of transformative innovation policies. Hence, when the evaluation of transformative innovation policies still rests on New Public Management principles, the shift towards transformative innovation policy will be hard unless it is followed by a shift in evaluation practices (Rohracher et al., 2023). This research, while specific to one ministry in one country, can still contribute to the broader field of transformative innovation policy with the government as a central actor in the execution and evaluation of these innovation policies.

A second limitation is methodological in nature. The analysis consisted of multiple consecutive steps. While this approach makes a strong case in step-by-step developing an evaluation approach and suggesting recommendations to move towards a learning-based evaluation framework, the steps limited the researchers in terms of data collection. The study focussed on perspectives from policy evaluation experts to propose a reflexive evaluation approach. Although these experts have extensive knowledge and expertise in the field of policy evaluation, the study leaves out the perspectives of those more closely involved in policy implementation and execution, such as policymakers and stakeholders. The reflexive approach requires the active involvement of these actors as the evaluation approach asks for continuous reflection of all policy actors to steer and improve policies. Hence, a shortcoming of this study resides in the exclusion of perspectives to whom the reflexive approach is applicable. Although the suggested approach encompasses stakeholder participation by including them in problem conceptualisations and developing a flexible Theory of Change, some pitfalls may arise. The empirical data gathered through reflexive evaluations depends not only on the development of a reflexive theoretical framework but relies heavily on the willingness of policy actors to share their knowledge and experiences in the course of the evaluation. To enrich the proposed approach, further empirical research could implement the approach in a real-life context. This could reveal the practical barriers to implementing a reflexive evaluation approach in an institutional context. Combining insights in other recently emerging reflexive evaluation approaches within Dutch Ministries (e.g. Folkert et al., 2020; Nabielek et al., 2023), a reflexive evaluation framework can be developed that encompasses both a theoretical and empirical understanding of learning-oriented evaluation practices.

The overall contributions of this study lie in addressing the evaluation challenges for TIPs. The discussion with participants together with emerging initiatives that implement new approaches for evaluating TIPs gives evidence that these developments are here to stay. Nevertheless, tackling

societal challenges will require developing and institutionalizing evaluation approaches suitable to the evaluation of TIPs. To do so, research and practice must narrow the gap between learning and accountability in evaluation practices. Although this research had its limitations, it is a first step in moving towards a culture of learning in policy development and evaluation.

References

- Abma, T. (1996). *Responsief evalueren*. Eburon Uitgeverij B.V.
- Abma, T., & in 't Veld, R. (2001). *Handboek Beleidswetenschap*. Amsterdam: Boom.
- Amanatidou, E., Cunningham, P., Gök, A., & Garefi, I. (2014). Using Evaluation Research as a Means for Policy Analysis in a 'New' Mission-Oriented Policy Context. *Minerva*, 52(4), 419–438.
<https://doi.org/10.1007/s11024-014-9258-x>
- Argyris, Ch., & Schön, D. A. (1997). Organizational Learning: A Theory of Action Perspective. *Reis*, 77/78, 345–348. <https://doi.org/10.2307/40183951>
- Arnold, E. (2004). Evaluating research and innovation policy: A systems world needs systems evaluations. *Research Evaluation*, 13(1), 3–17. <https://doi.org/10.3152/147154404781776509>
- Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors* (pp. 609–626). Princeton University Press. <https://www.nber.org/books-and-chapters/rate-and-direction-inventive-activity-economic-and-social-factors/economic-welfare-and-allocation-resources-invention>
- Audouin, M., & Finger, M. (2018). The development of Mobility-as-a-Service in the Helsinki metropolitan area: A multi-level governance analysis. *Research in Transportation Business & Management*, 27, 24–35. <https://doi.org/10.1016/j.rtbm.2018.09.001>
- Banner, D. K. (1974). The politics of evaluation research. *Omega*, 2(6), 763–774.
[https://doi.org/10.1016/0305-0483\(74\)90115-7](https://doi.org/10.1016/0305-0483(74)90115-7)
- Behn, R. D. (1998). The new public management paradigm and the search for democratic accountability. *International Public Management Journal*, 1(2), 131–164. [https://doi.org/10.1016/S1096-7494\(99\)80088-9](https://doi.org/10.1016/S1096-7494(99)80088-9)
- Bergek, A., Hekkert, M., Jacobsson, S., Markard, J., Sandén, B., & Truffer, B. (2015). Technological innovation systems in contexts: Conceptualizing contextual structures and interaction dynamics. *Environmental Innovation and Societal Transitions*, 16, 51–64.
<https://doi.org/10.1016/j.eist.2015.07.003>
- Bergek, A., Jacobsson, S., Carlsson, B., Lindmark, S., & Rickne, A. (2008). Analyzing the functional dynamics of technological innovation systems: A scheme of analysis. *Research Policy*, 37(3), 407–429.
<https://doi.org/10.1016/j.respol.2007.12.003>

- Borrás, S., & Edler, J. (2020). The roles of the state in the governance of socio-technical systems' transformation. *Research Policy*, 49(5), 103971. <https://doi.org/10.1016/j.respol.2020.103971>
- Braams, R. B., Wesseling, J. H., Meijer, A. J., & Hekkert, M. P. (2021). Legitimizing transformative government: Aligning essential government tasks from transition literature with normative arguments about legitimacy from Public Administration traditions. *Environmental Innovation and Societal Transitions*, 39, 191–205. <https://doi.org/10.1016/j.eist.2021.04.004>
- Braams, R. B., Wesseling, J. H., Meijer, A. J., & Hekkert, M. P. (2023). Civil servant tactics for realizing transition tasks understanding the microdynamics of transformative government. *Public Administration*, n/a(n/a). <https://doi.org/10.1111/padm.12933>
- Caniëls, M. C. J., & Romijn, H. A. (2008). Strategic niche management: Towards a policy tool for sustainable development. *Technology Analysis & Strategic Management*, 20(2), 245–266. <https://doi.org/10.1080/09537320701711264>
- Carlsson, B., & Stankiewicz, R. (1991). On the nature, function and composition of technological systems. *Journal of Evolutionary Economics*, 1(2), 93–118. <https://doi.org/10.1007/BF01224915>
- Chiang, J.-T. (1991). From 'mission-oriented' to 'diffusion-oriented' paradigm: The new trend of U.S. industrial technology policy. *Technovation*, 11(6), 339–356. [https://doi.org/10.1016/0166-4972\(91\)90017-X](https://doi.org/10.1016/0166-4972(91)90017-X)
- Cooke, P., Gomez Uranga, M., & Etxebarria, G. (1997). Regional innovation systems: Institutional and organisational dimensions. *Research Policy*, 26(4), 475–491. [https://doi.org/10.1016/S0048-7333\(97\)00025-5](https://doi.org/10.1016/S0048-7333(97)00025-5)
- Daimer, S., Hufnagl, M., & Warnke, P. (2012). Challenge-oriented policy-making and innovation systems theory. *Fraunhofer ISI (Hg.): Innovation System Revisited*, 40, 217–234.
- Diercks, G., Larsen, H., & Steward, F. (2019). Transformative innovation policy: Addressing variety in an emerging policy paradigm. *Research Policy*, 48(4), 880–894. <https://doi.org/10.1016/j.respol.2018.10.028>
- Ergas, H. (1986). *Does Technology Policy Matter?* (SSRN Scholarly Paper No. 1428246). <https://doi.org/10.2139/ssrn.1428246>
- Folkert, R., Bouwma, I., Kuindersma, W., van der Hoek, D. J., Gerritsen, A., & Kunseler, E. (2020). *Lerende evaluatie van het Natuurpact 2020: Gezamenlijk de puzzel leggen voor natuur, economie en maatschappij: Tweede rapportage* (No. 3852). PBL Planbureau voor de Leefomgeving.

- Geels, F., & Raven, R. (2006). Non-linearity and Expectations in Niche-Development Trajectories: Ups and Downs in Dutch Biogas Development (1973–2003). *Technology Analysis & Strategic Management*, 18(3–4), 375–392. <https://doi.org/10.1080/09537320600777143>
- Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: A multi-level perspective and a case-study. *Research Policy*, 31(8), 1257–1274. [https://doi.org/10.1016/S0048-7333\(02\)00062-8](https://doi.org/10.1016/S0048-7333(02)00062-8)
- Ghosh, B., Kivimaa, P., Ramirez, M., Schot, J., & Torrens, J. (2021). Transformative outcomes: Assessing and reorienting experimentation with transformative innovation policy. *Science and Public Policy*, 48(5), 739–756. <https://doi.org/10.1093/scipol/scab045>
- 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. <https://doi.org/10.1016/j.respol.2018.10.004>
- Grin, J., Felix, F., Bos, B., & Spoelstra, S. (2004). Practices for reflexive design: Lessons from a Dutch programme on sustainable agriculture. *International Journal of Foresight and Innovation Policy*, 1(1–2), 126–149. <https://doi.org/10.1504/IJFIP.2004.004618>
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. SAGE.
- Haddad, C. R., Nakić, V., Bergek, A., & Hellsmark, H. (2022). Transformative innovation policy: A systematic review. *Environmental Innovation and Societal Transitions*, 43, 14–40. <https://doi.org/10.1016/j.eist.2022.03.002>
- Haddad, R. (2021). Evaluating transformative innovation policy: Towards an integrated framework. *Sweden: (Doctoral Dissertation, Chalmers Tekniska Hogskola)*.
- 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. <https://doi.org/10.1016/j.eist.2019.11.011>
- Hekkert, M. P., Suurs, R. A. A., Negro, S. O., Kuhlmann, S., & Smits, R. E. H. M. (2007). Functions of innovation systems: A new approach for analysing technological change. *Technological Forecasting and Social Change*, 74(4), 413–432. <https://doi.org/10.1016/j.techfore.2006.03.002>
- Hoogerwerf, A. (1984). *Beleid berust op veronderstellingen: De beleidstheorie*.
- I&W. (2016). *Smart Mobility: Bouwen aan een nieuw tijdperk op onze wegen*. Den Haag: Ministry of Infrastructure and Water Management.

- I&W. (2021). *Duurzame mobiliteit*. Magazines Rijksoverheid.
<https://magazines.rijksoverheid.nl/ienw/duurzaamheidsverslag/2021/01/duurzame-mobiliteit>
- 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.
<https://doi.org/10.1016/j.techfore.2018.08.011>
- Janssen, M. J., Bergek, A., & Wesseling, J. H. (2022). Evaluating systemic innovation and transition programmes: Towards a culture of learning. *PLOS Sustainability and Transformation*, *1*(3), e0000008. <https://doi.org/10.1371/journal.pstr.0000008>
- Kallerud, E., Amanatidou, E., Upham, P., Nieminen, M., Klitkou, A., Olsen, D., Lima Toivanen, M., Oksanen, J., & Scordato, L. (2013). *Dimensions of Research and Innovation Policies to Address Grand and Global Challenges*.
- Kanger, L., Sovacool, B. K., & Noorköiv, M. (2020). Six policy intervention points for sustainability transitions: A conceptual framework and a systematic literature review. *Research Policy*, *49*(7), 104072. <https://doi.org/10.1016/j.respol.2020.104072>
- Kattel, R., & Mazzucato, M. (2018). Mission-oriented innovation policy and dynamic capabilities in the public sector. *Industrial and Corporate Change*, *27*(5), 787–801.
<https://doi.org/10.1093/icc/dty032>
- Kemp, R., Schot, J., & Hoogma, R. (1998). Regime shifts to sustainability through processes of niche formation: The approach of strategic niche management. *Technology Analysis & Strategic Management*, *10*(2), 175–198. <https://doi.org/10.1080/09537329808524310>
- Keulen, S. (2020). 20 jaar Verantwoordingsdag: Inzicht voor Kamercommissies. *Bestuurskunde*, *20*(1), 64–75. <https://doi.org/10.5553/Bk/092733872020029001008>
- Kivimaa, P., & Kern, F. (2016). Creative destruction or mere niche support? Innovation policy mixes for sustainability transitions. *Research Policy*, *45*(1), 205–217.
<https://doi.org/10.1016/j.respol.2015.09.008>
- Köhler, J., Geels, F. W., Kern, F., Markard, J., Onsongo, E., Wieczorek, A., Alkemade, F., Avelino, F., Bergek, A., Boons, F., Fünfschilling, L., Hess, D., Holtz, G., Hyysalo, S., Jenkins, K., Kivimaa, P., Martiskainen, M., McMeekin, A., Mühlemeier, M. S., ... Wells, P. (2019). An agenda for sustainability transitions research: State of the art and future directions. *Environmental Innovation and Societal Transitions*, *31*, 1–32. <https://doi.org/10.1016/j.eist.2019.01.004>

- Kuhlmann, S., & Rip, A. (2018). Next-Generation Innovation Policy and Grand Challenges. *Science and Public Policy*, 45(4), 448–454. <https://doi.org/10.1093/scipol/scy011>
- Lazarevic, D., Salo, H., & Kautto, P. (2022). Circular economy policies and their transformative outcomes: The transformative intent of Finland's strategic policy programme. *Journal of Cleaner Production*, 379, 134892. <https://doi.org/10.1016/j.jclepro.2022.134892>
- Lee, Y.-N. (2015). Evaluating and extending innovation indicators for innovation policy. *Research Evaluation*, 24(4), 471–488. <https://doi.org/10.1093/reseval/rvv017>
- Loorbach, D. (2007). *Transition Management: New mode of governance for sustainable development*. <https://repub.eur.nl/pub/10200/>
- Loorbach, D. (2010). Transition Management for Sustainable Development: A Prescriptive, Complexity-Based Governance Framework. *Governance*, 23(1), 161–183. <https://doi.org/10.1111/j.1468-0491.2009.01471.x>
- Lundvall, B. A. (1992). *National systems of innovation: Towards a theory of innovation and interactive learning*.
- Magro, E., & Wilson, J. R. (2019). Policy-mix evaluation: Governance challenges from new place-based innovation policies. *Research Policy*, 48(10), 103612. <https://doi.org/10.1016/j.respol.2018.06.010>
- Malerba, F. (2002). Sectoral systems of innovation and production. *Research Policy*, 31(2), 247–264. [https://doi.org/10.1016/S0048-7333\(01\)00139-1](https://doi.org/10.1016/S0048-7333(01)00139-1)
- Mazzucato, M. (2016). From market fixing to market-creating: A new framework for innovation policy. *Industry and Innovation*, 23(2), 140–156. <https://doi.org/10.1080/13662716.2016.1146124>
- 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. <https://doi.org/10.1093/reseval/rvab016>
- Nabielek, P., Vandenbussche, L., Loeber, A., Boonstra, H., Klaassen, P., & Verwoerd, L. (2023). *WORKING PAPER: AN EVALUATION FRAMEWORK FOR THE TRANSFORMATIVE CAPACITY OF DUTCH CLIMATE POLICY*. PBL Netherlands Environmental Assessment Agency.
- Negro, S. O., Suurs, R. A. A., & Hekkert, M. P. (2008). The bumpy road of biomass gasification in the Netherlands: Explaining the rise and fall of an emerging innovation system. *Technological Forecasting and Social Change*, 75(1), 57–77. <https://doi.org/10.1016/j.techfore.2006.08.006>

- Norris, N., & Kushner, S. (2007). The New Public Management and Evaluation. In S. Kushner & N. Norris (Eds.), *Dilemmas of Engagement: Evaluation and the New Public Management* (Vol. 10, pp. 1–16). Emerald Group Publishing Limited. [https://doi.org/10.1016/S1474-7863\(07\)10001-6](https://doi.org/10.1016/S1474-7863(07)10001-6)
- OECD. (2020). *Improving Governance with Policy Evaluation: Lessons From Country Experiences*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/governance/improving-governance-with-policy-evaluation_89b1577d-en
- Parliamentary letter, no. 31305, 260 (2018).
- Parliamentary letter, no. 31305, 294 (2019).
- Parliamentary letter, no. 31505, 271 (2018).
- Pattyn, V., & Verweij, S. (2014). *Beleidsevaluaties Tussen Methode En Praktijk: Naar Een Meer Realistische Evaluatiebenadering (Policy Evaluations between Method and Practice: Towards a More Realistic Evaluation Approach)* (SSRN Scholarly Paper No. 2551506). <https://papers.ssrn.com/abstract=2551506>
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage.
- Quist, J. (2007). *Backcasting for a Sustainable Future: The Impact After 10 Years*. Eburon Uitgeverij B.V.
- Reinertsen, H., Bjørkdahl, K., & McNeill, D. (2022). Accountability versus learning in aid evaluation: A practice-oriented exploration of persistent dilemmas. *Evaluation*, 28(3), 356–378. <https://doi.org/10.1177/13563890221100848>
- Rip, A., & Kemp, R. (1998). Technological change. *Human Choice and Climate Change*, 2(2), 327–399.
- Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4(2), 155–169. <https://doi.org/10.1007/BF01405730>
- Rohracher, H., Coenen, L., & Kordas, O. (2023). Mission incomplete: Layered practices of monitoring and evaluation in Swedish transformative innovation policy. *Science and Public Policy*, 50(2), 336–349. <https://doi.org/10.1093/scipol/scac071>
- Rotmans, J., Kemp, R., & van Asselt, M. (2001). More evolution than revolution: Transition management in public policy. *Foresight*, 3(1), 15–31. <https://doi.org/10.1108/14636680110803003>
- Schlaile, M. P., Urmetzer, S., Blok, V., Andersen, A. D., Timmermans, J., Mueller, M., Fagerberg, J., & Pyka, A. (2017). Innovation Systems for Transformations towards Sustainability? Taking the Normative Dimension Seriously. *Sustainability*, 9(12), Article 12. <https://doi.org/10.3390/su9122253>

- Schot, J., & Geels, F. W. (2008). Strategic niche management and sustainable innovation journeys: Theory, findings, research agenda, and policy. *Technology Analysis & Strategic Management*, 20(5), 537–554. <https://doi.org/10.1080/09537320802292651>
- Schot, J., Kivimaa, P., & Torrans, J. (2019). *Transforming experimentation: Experimental policy engagements and their transformative outcomes. (TIPC Research Report, March)*. <https://doi.org/10.19710>
- Schot, J., & Steinmueller, W. E. (2018). Three frames for innovation policy: R&D, systems of innovation and transformative change. *Research Policy*, 47(9), 1554–1567. <https://doi.org/10.1016/j.respol.2018.08.011>
- Smith, A., Voß, J.-P., & Grin, J. (2010). Innovation studies and sustainability transitions: The allure of the multi-level perspective and its challenges. *Research Policy*, 39(4), 435–448. <https://doi.org/10.1016/j.respol.2010.01.023>
- Snellen, D., Bastiaanssen, J., & 't Hoen, M. (2021). *Brede welvaart en mobiliteit*. Den Haag: PBL.
- Suurs, R. A. A., & Hekkert, M. P. (2009). Cumulative causation in the formation of a technological innovation system: The case of biofuels in the Netherlands. *Technological Forecasting and Social Change*, 76(8), 1003–1020. <https://doi.org/10.1016/j.techfore.2009.03.002>
- TwynstraGudde. (2022). *Evaluatie Programma MaaS*. <https://open.overheid.nl/documenten/ronl-55a177663787c551e1f2595bd5fcf755da12bc4e/pdf>
- Van der Knaap, P. (1995). Policy Evaluation and Learning: Feedback, Enlightenment or Argumentation? *Evaluation*, 1(2), 189–216. <https://doi.org/10.1177/135638909500100205>
- van der Knaap, P., Pattyn, V., & Hanemaayer, D. (2020). *Beleidsevaluatie in theorie en praktijk*. <https://lirias.kuleuven.be/3180021>
- van Mierlo, B., Arkesteijn, M., & Leeuwis, C. (2010). Enhancing the Reflexivity of System Innovation Projects With System Analyses. *American Journal of Evaluation*, 31(2), 143–161. <https://doi.org/10.1177/1098214010366046>
- Voß, J.-P., & Kemp, R. (2015). *Sustainability and reflexive governance: Introduction*. Technische Universität Berlin.
- Weber, K. M., & 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(6), 1037–1047. <https://doi.org/10.1016/j.respol.2011.10.015>

- Weiss, C. H. (1998). Have We Learned Anything New About the Use of Evaluation? *American Journal of Evaluation*, 19(1), 21-33. <https://doi.org/10.1177/109821409801900103>
- Wildavsky, A. B. (1979). *Speaking truth to power*. Transaction Publishers.
- Woolthuis, R. K., Lankhuizen, M., & Gilsing, V. (2005). A system failure framework for innovation policy design. *Technovation*, 25(6), 609-619. <https://doi.org/10.1016/j.technovation.2003.11.002>

Appendix A: Reviewed chapters of the annual budget of I&W

Chapter	Title
B.2.1.	Beleidsprioriteiten
B.3.3.	Wegen en verkeersveiligheid
B.3.4.	Openbaarvervoer en Spoor
B.3.5.	Luchtvaart
B.3.6.	Scheepvaart en havens
B.3.7.	Uitvoering milieubeleid en internationaal
B.3.8.	Lucht en geluid

Appendix B: Codebook (Step 1 in methods)

Code name	Description
Destabilize the unsustainable	
Control policies and make significant changes in regime rules	Introduce extra goals and measures to redirect adverse developments, reform tax system to tax the unsustainable, restrict use of unsustainable practices, introduce policies that erode unsustainable regimes.
Reduce support for dominant regime technologies	Address market failures responsible for unsustainability, provide evidence from experiments for regime shifts, slow down or stop new unsustainable developments.
Develop internal capabilities and structures	
Development of new competencies	Become more entrepreneurial, analyse innovation systems, build dynamic organizational capabilities, understand new technological developments.
Establish mechanisms for policy coordination	Coordinate between public institutes, create new institutional conditions, embed processes in institutes, set up responsible institutes.
Monitor and evaluate	Continuous monitoring and evaluation, develop the capacity for learning, learn to experiment and explore.
Rethink own role in a transition	Take a holistic perspective, align social and environmental challenges with national innovation objectives, embrace opportunities, internal focus on upscaling, revise and critically evaluate own role and regulation.
Give direction	
Articulate direction	Articulate demand, develop missions, guiding role and show leadership in structural change, state ambition and set targets, select experiments, translate ideas into priorities and actions, create a vision for the future
Construct policy strategies in order to direct	Create public organizations to link emerging markets with societal challenges, create stable policy frameworks regarding guidance and market formation, justify new policies and government intervention.
Direct through enforced regulations	Enforce laws and IP rights, standardize and regulate.
Reconfigure the market	Create and shape markets, form markets through minimal consumption quotas, give direction through establishing a favourable tax regime, give legitimacy to a technological field, help the market decide on strategic investments.
Support governance	
Activate actors	Acknowledge the third sector and consumers, encourage parties to participate, make room for a variety of voices, arguments and interpretations.
Goals achieving strategies	Ensure the process of co-evolution leads to a desirable outcome, facilitate reciprocal learning from experimentation, mobilize private

	financial organizations, organize platforms for collective action, stimulate collective learning process, stimulate discussion.
Guiding organizational arrangements	Create coalitions and make covenants, facilitate the development of networks, facilitate Public-Private Partnerships, improve governance, mediate in brokering, be the niche manager
Support the new	
Engage in entrepreneurial experiments	Embrace innovation as an option and make it assessable, engage with new niche actors, organize interaction between emergent technology groups and government, steer from within a niche, provide room for experimentation.
Establish market formation	Build beneficial infrastructure for innovations, Create, protect and facilitate niches, give temporary exemption from regulations, mitigate initial negative impact of innovation, remove institutional barriers, stimulate and initiate new pilots and developments, support diffusion.
Help new developments develop and diffuse	Introduce and demonstrate new technologies and use them to set expectations, communicate about new developments, develop sufficient technological variation, train third parties' capacity and capability.
Price-performance improvement and resource mobilization	Create innovation funds, fund education, fund experiments, invest in new technologies, public procurement, stimulate with materials and subsidies, support complementary technologies, support research, help find funding.

Appendix C: Interviewee participants

Interviewee number*	Interviewee description
1	Senior consultant who evaluated transformative policy programs
2	Consultant who evaluated transformative policy programs
3	Associate professor who participated in an expert committee in evaluating transition policies
4	Senior policy advisor part of an advice counsel on climate policy issues
5	Assistant professor and scientist in innovation and transition policy who participated in an expert committee in evaluating transition policies
6	Environmental policy scientist at the Dutch Environmental Assessment Agency specified in policy learning in transitions and reflexive evaluations
7	Associate professor in governance and sustainability involved in the evaluation of multiple transformative programmes
8	Senior researcher at the Dutch Environmental Assessment Agency who reflected on transformative policy evaluations in mobility
9	Researcher at the Rathenau Institute specialized in the contribution of science and innovation policy to grand challenges
10	Coordinating advisor at the directorate of financial economic affairs who coordinated multiple learning-based pilots within a Ministry

*The order of interview participants is randomized and do not follow in text references

Appendix D: Interview guide

Introduction:

You are invited to take part in this study on an evaluation approach for policies aiming for transformative change. The study is conducted by Sander van de Wijngaert who is a student in the MSc programme Innovation Sciences at the Department of Sustainable Development, Utrecht University, and part of the team Innovation in Mobility at the Ministry of Infrastructure and Water Management (I&W).

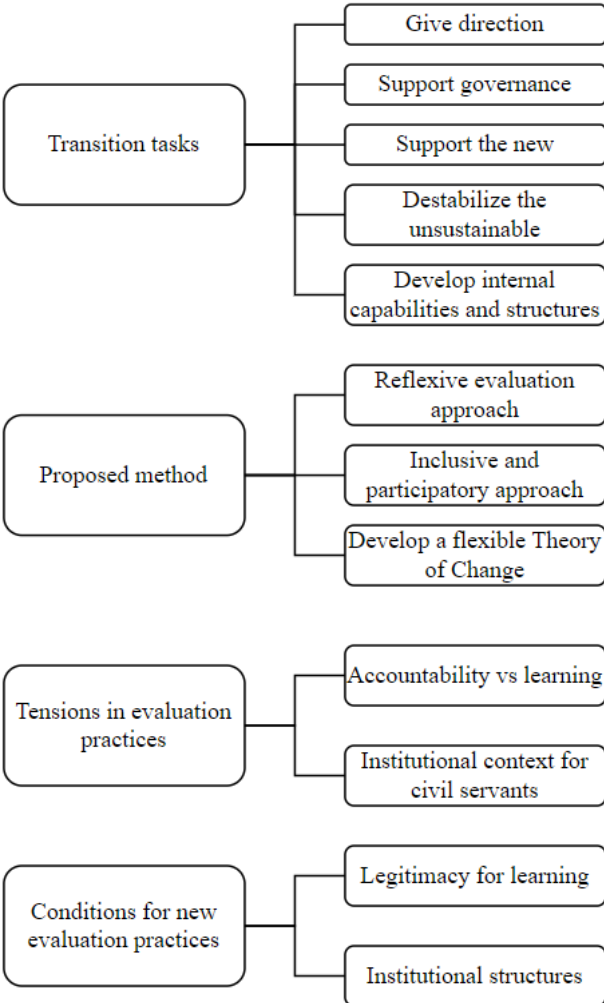
Your participation in this interview is completely voluntary. You can quit at any time without providing any reason and without any penalty. Your contribution to the study is valuable to us, and we appreciate your time to complete this interview. We estimate it will take about 60 minutes to complete the interview. The questions will be read out to you by the interviewer. Some of the questions require little time to complete, while other questions might need more careful consideration. Please feel free to skip questions you do not feel comfortable answering. You can also ask the interviewer to clarify or explain questions you find unclear before providing an answer. Your answers will be noted by the interviewer in an answer template. The data you provide will be used for writing a Master thesis report and may be used for other scientific purposes such as a publication in a scientific journal or presentation at academic conferences.

The interview will be audio taped with your consent for transcription purposes. The audio recordings will be available to the Master student and supervisors. We will process your data confidentially and in accordance with data protection legislation (the General Data Protection Regulation and Personal Data Act). Audio recordings will only be stored on a secured and encrypted server of Utrecht University.

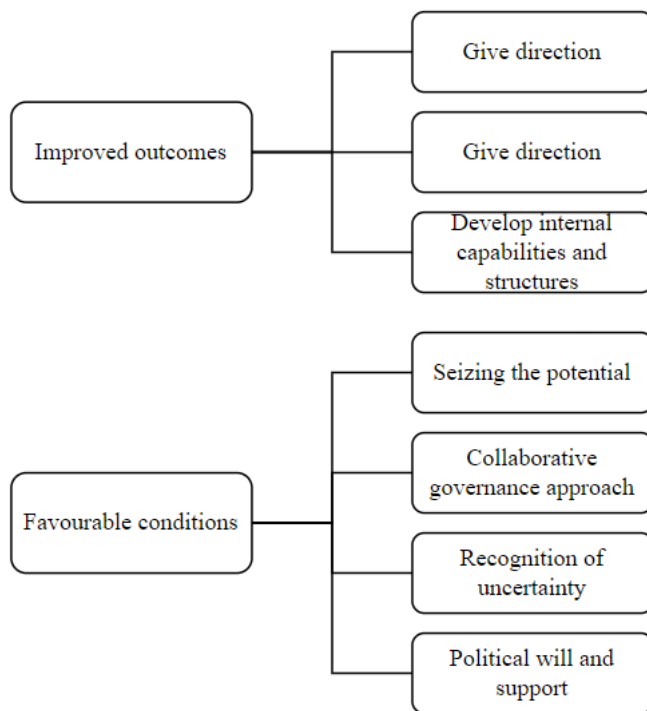
Interview questions:

1. What are your experiences and perspectives on evaluating policies that aim for transitions?
2. Based on your experiences, what are the challenges for evaluating transition policies and how can we address these?
3. What are effective frameworks that you think suit the evaluation of transition policy?
4. [Researcher discusses transition tasks] Which transition tasks do you recognise in policy evaluation?
5. How do you think these tasks could be operationalized in an evaluation approach?
6. How can we make sure this evaluation approaches can be institutionalized in evaluation practices?
7. Which tensions do you see in how evaluations should take place and how evaluations are conducted?
8. What is necessary to shift towards new evaluation practices?
9. What question should I have asked you that we have not yet covered?

Appendix E: Coding tree (Step 3 in methods)



Appendix F: Coding tree (Step 4 in methods)



Appendix G: description of abbreviations in Table 5

Abbreviation	Description	English translation
MaaS	Programma Mobility as a Service	Mobility as a Service program
NAL	Nationale Agenda Laadinfrastructuur	national agenda for charging infrastructure
TBOV	Toekomstbeeld Openbaar Vervoer	future vision on public transport
US	Uitvoeringsagenda Stadslogistiek	implementation agenda of city logistics
DKTI	Demonstratieregeling Klimaat Technologieën en -Innovaties in transport	demonstration scheme for climate technologies and innovations in transport
DBL	Duurzame Brandstoffen Luchtvaart	sustainable fuels for aviation
CETSI V	Rijkscofinancieringsregeling V	government co-financing scheme V
SIDB	Subsidieregeling Innovaties Duurzame Binnenvaart	subsidy scheme for innovations in sustainable inland shipping
SGV	Maatregelenpakket Spoorgoederenvervoer	measures for rail freight transport