EVALUATING POLICY DESIGN SPRINT OUTCOMES

A Proposition for a Novel Tool

Jānis Ķesa

Master's Thesis

Service Design Strategies and Innovations (SDSI)

University of Lapland

Spring 2025

ABSTRACT

Policy Design Sprints are increasingly used in public policy development, but their long-term impacts are often unclear due to the lack of systematic evaluation frameworks. This thesis addresses this gap by developing and testing a structured evaluation tool (the "Better Outcome" tool) for sprint outcomes. Drawing on interdisciplinary literature (Service Design, Innovation Studies, Legal Design, Systems and Futures Thinking, and Behavioral Economics) and government practice, the tool evaluates outcomes against multiple criteria, including implementation feasibility, systemic alignment, behavioral uptake, and future adaptability, to determine their viability and value. Using a qualitative case-study approach with two pilot sprints in Latvian government, data from participant surveys, interviews with practitioners, and expert walkthroughs were collected and used to iteratively refine the tool. The empirical findings confirm a persistent "last-mile" gap: many promising ideas stall after the sprint. Applying the framework helped participants clarify and score proposed solutions on the defined dimensions, facilitating discussion of strengths and weaknesses. Stakeholders reported that the tool improved clarity in outcome assessment and alignment with institutional objectives. Overall, the research demonstrates that a multidimensional, criteria-based evaluation framework can bridge creative ideation and policy implementation. The Better Outcome tool offers policymakers an actionable method to rigorously assess and guide design-led policy interventions, thereby enhancing accountability, evidence-based decision-making, and the sustainable impact of public-sector innovation.

Keywords: *Policy Design, Design Sprints, Public Sector Innovation, Outcome Evaluation, Systems Thinking.*

Supervisor: *Nam Kiwoong*

Number of pages:*100 (one hundred)*

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

1.1 Background of the Research

In the past decade, public sector and governmental organisations have been facing greater challenges that require new, innovative, collaborative, and human-centred approaches to policymaking. Global crises (for example, climate change and pandemics), as well as growing demands for adaptive governance models and responsiveness to societal issues, have compounded the challenges in the public sector. This has fueled new public governance settings focusing on co-creation, participatory policymaking, and Service Design methods (Bason, 2018).

Service Design, in the public service context, revolves around the design of services following principles of design thinking and systems theory by improving services through user research, co-creation (co-design) and rapid prototyping (Stickdorn et al., 2018). Service Design (or Public Service Design) in the public sector means developing services that are more efficient, equitable, and human-centred, which helps improve citizen satisfaction with government services and organisational responsiveness (Holmlid & Evenson, 2008). Similarly, innovation studies help researchers explore how new ideas, approaches, and technologies diffuse through systems, providing frameworks for understanding social change in bureaucratic spaces and multi-stakeholder ecosystems (Mulgan & Albury, 2003).

Legal Design is concerned with designing legal and regulatory systems with design principles to create more user-friendly and accessible legal and regulatory systems (Haapio & Hagan, 2021).

Design Sprints are a practice-based, structured five-phase process for quickly addressing problems and testing ideas. They represent an example of design within government-related contexts. The Design Sprint framework was developed by Knapp et al. (2016) for Google Ventures and was later adapted for broader use, including public policymaking. Prominent pioneering institutions (for example, UK Policy Lab) and the OECD (UK Policy Lab, 2021; OECD, 2021) have adapted Policy Design Sprints based on the above methodology to find the potential value of this approach for tackling complex policy contexts involving co-creation, stakeholder engagement, and iterative development.

Although their popularity is growing in Latvia, the ability to assess the success of Policy Design Sprints is still largely underdeveloped. There are existing evaluation methods in the legislative process in Latvia (Gailīte, 2023), but structured formats specifically for the outputs of public Service Design sprints are limited. Conventional policy evaluation methods do not tend to address the qualitative and iterative nature of designing approaches to public service, particularly through highly time-constrained Design Sprints. This is why it is imperative that evaluation occurs through expanded methodologies that synthesise lessons from Service Design, Innovation Studies, and Legal Design.

This thesis addresses this knowledge gap by advocating a structured evaluation framework for Policy Design Sprints that reflects Systems Thinking, behavioural insights, and foresight methodologies. The research is grounded in contemporary Service Designs, which emphasise user engagement as well as multi-stakeholder co-creation. It draws from Innovation Studies in terms of institutional change and the adoption of new practices and uses Legal Design to understand the regulatory nature of policy interventions.

The relevance of this research is based on its ability to offer a replicable framework for evaluating the output of design-initiated policy projects. As governments explore more agile and human-centred methods, it is important to place emphasis on assessing these approaches systematically. This study bridges theoretical knowledge and practical tools: it advances the academic conversation and practical support for facilitators as well as policymakers and public administrators. Ultimately, by establishing a foundation for structured evaluation, the research encourages evidence-based decision-making and an ongoing developmental path toward improved design-led public innovation.

1.2 Research motivation

Policy Design Sprints have become increasingly popular with Latvian government institutions as a rapid, collaborative method to tackle complex problems. Sprints offer a formatted, bounded way to leverage cross-disciplinary idea generation, stakeholder engagement, and rapid prototyping, all of which are the least evolved components of a bureaucratic policy process. Despite experiences of effective processing and collaborative work, the long-term implications of sprint outcomes are typically limited or incomplete.

The disconnect between energised co-creation and limited implementation creates a challenging problem. The effectiveness of Policy Design Sprints is unclear — particularly

whether ideas generated in a sprint impact the construction of meaningful, implementable policy innovations. The closures of recently created public innovation labs or design units often reveal that without demonstrated impact, support for this activity is short-lived (Timeus & Gascó, 2018b; Monteiro & Kumpf, 2023). Ordinary ways of assessing the success of design approaches via traditional metrics (e.g., budget spent, number of ideas generated) provide only a shallow notion of their potential value since process metrics do not typically reveal if the sprint produced an implementable impactful policy.

The author is motivated by his firsthand personal experiences of facilitating Policy Design Sprints. Specifically, in government settings, it was observed that many sprint participants leave sprint sessions energised and aligned, but organisations often lack the institutional capacity to implement the ideas forwarded from the sprint. This observation provoked the search for evaluation metrics that assess outcomes beyond outputs, which are likely performance-based. The proposed tool assesses aspects such as:

- Alignment with original sprint objectives and problem statements.
- Future resilience and adaptability (*Futures Thinking*).
- Systemic coherence and long-term impact (*Systems Thinking*).
- behavioural effectiveness and policy adoption (*behavoural Economics*).

The tool assesses whether the outcomes of the Policy Design Sprint meets the criteria for the sprint goals, but also whether the ideas are forward-compatible and systemically viable. The tool seeks to present sprint teams and decision-makers with a multi-dimensional perspective of the outcomes, incorporating a futures analysis and behavoural perspective.

This research aims to bridge the gap between process and outcome evaluation in Policy Design Sprints. The goal was to enhance the robustness of design-based policymaking, addressing the need for greater methodological rigour. Additionally, the study sought to enrich the knowledge base related to Service Design by illustrating how design processes can help mitigate the limitations faced by the public sector. This research is well timed as governments increasingly depend on agile methods to respond to critical issues (e.g., climate action, digital transformation), and there is growing scepticism regarding the effectiveness of, and legitimacy of these methods (Mergel et al., 2018; Neumann et al., 2024). An evaluation mechanism provides a route for governments to moderate the

experimental nature of the design method and research, as well as the accountability demands of public governance.

The key ambitions of this thesis is to facilitate the legitimacy and effectiveness of Policy Design Sprints through structured outcome evaluation, and to contribute to academic debate and knowledge generation by integrating frameworks from Service Design, Innovation, and Legal Design. Ultimately, it would contribute to facilitating public sector entities to better assess and operationalise the outputs from design-driven policy processes.

1.3 Research goals and research questions

This research aims to develop a logical, multi-dimensional framework to assess the outcomes of Policy Design Sprints in governments. There is an increased interest in sprint methodology in the public service, but there are also challenges in converting sprint outputs into policy action, highlighting the need for a framework to measure outcomes. Though Policy Design Sprints are effective for generating ideas, engaging stakeholders, developing provisional solutions, and having a long-term impact, the lack of outcome measurement undermines their long-term value. The author seeks to bridge the gap in the literature with an interdisciplinary approach informed by Service Design, Innovation Studies, and Legal Design.

The proposed solution or the tool is designed to measure sprint outcomes by defined criteria like implementation viability, contribution to overarching policy goals, social equity, democratic participation, systemic coherence, and adaptability to future challenges. The research revolves around the author's professional experience facilitating Policy Design Sprints in response to the pressing need for evaluation methodologies that reach beyond the use of classical metrics toward outcome-oriented values reflecting the complex, multi-dimensional nature of policymaking.

The specific objectives of this research are as follows:

1. To explore existing literature and practices related to the evaluation of Policy Design Sprints within the contexts of Service Design, Innovation Studies, and Legal Design.
2. To identify key success criteria for evaluating sprint outcomes, including implementability, alignment with policy goals, societal impact, systemic coherence,

behavoural relevance, and future-readiness through interviews, observations, and document analysis.

3. To design a tool for evaluation of Policy Design Sprint outcome - a tool that operationalizes these criteria into a practical, replicable methodology suitable for use by facilitators operating in public sector co-creation workshops.
4. To assess the applicability and adaptability of the evaluation tool and propose refinements based on expert opinions.
5. To contribute to the academic discourse on policy innovation by integrating multiple disciplinary perspectives into a unified evaluation approach.

The central research question guiding this study is:

How can the outcomes of Policy Design Sprints in government be effectively evaluated in terms of implementation viability, systemic relevance, behavoural alignment, and future adaptability?

Sub-questions include:

- What are the critical success factors that determine whether a policy sprint outcome is considered to be a success?
- How can evaluation criteria be derived from multidisciplinary frameworks, including Service Design, Innovation Studies, Futures Thinking, Systems Thinking, behavoural Economics, and Legal Design?
- What methodological approaches are most suitable for capturing both qualitative and systemic dimensions of sprint outcomes?
- How does the application of an evaluation tool influence stakeholder understanding, policy adoption, and institutional learning?

This research is important from theoretical perspectivas as well as from the practical view. As governmental institutions become more agile and more accountable to citizens, they require strong mechanisms to assess the effectiveness of the working processes within. Without monitoring, it is too easy to consider the newest and most progressive approaches to public policy ineffective, or not well-executed. This thesis has developed a structured, interdisciplinary, outcome-oriented evaluation framework which productively intersects learning by doing and evidence-based decision-making.

The research also indicates broader shifts in public governance. Governments continue to transition from traditional top-down governance to networked and participatory models, which allow for design-thinking processes such as Policy Design Sprints (Blomkamp, 2021b). Ultimately, for these mechanisms to become strong conventional tools, a way of demonstrating accountability and the value of achieving plausible public outcomes is needed. This research emphasises domain-specific significant dimensions of evaluations, which include user-centredness and iterative learning (Service Design), institutional diffusion (Innovation Studies), and clearly operationalisable (Legal Design).

In conclusion, this research aims to better prepare public institutions for implementation of Policy Design Sprints and provides a systematically robust evaluation framework for measuring policy outputs. The main objective is to change the focus from quality of process to quality of policy, from creative engagement to real sustainability, and shift from symbolic participation to tangible transformation. Through theoretical integration, methodological innovation, and empirical validation, this thesis strives to make a meaningful contribution to the evolution of design-led policymaking.

1.4 Research scope

This research sits across Service Design, Innovation Studies, and Legal Design. Given the paper's distinct focus on examining the outcomes of Policy Design Sprints undertaken in governance contexts, the research is both conceptual and empirical and tries to contribute to the theory of policy design and provide a tool to facilitate the movement of Policy Design Sprint facilitators and public service institutions to review and improve their sprint outcomes. The main component of the research piece is the development and use of a structured evaluation tool, **the "Better Outcome" tool**, to be used at the final evaluation step of the Policy Design Sprint. The tool addresses a particular gap in practice, the lack of rigorous, meaningful, and contextualised methods to evaluate the outcomes of sprints, in particular the ease of implementation of the sprint outcomes, connectedness to a system, relatable to behaviours, and readiness for the future.

Methodologically, the research uses a qualitative, exploratory approach using sequential mixed methods in order to achieve validity and to ensure it is relevant. The research began with a survey sent to participants of Policy Design Sprints in order to obtain broad findings. Next, semi-structured interviews with Policy Design Sprint team leaders provided

a deeper understanding built on richly described cases explained by the participants. In addition, the expert's walkthroughs of the prototype were used to test and refine the emerging tool through real-life experience, ensuring usability and relevance to practice. Each stage of the research is built on the previous stage, developing an iterative triangulated research design. This multi-faceted research design provided breadth, depth and expert knowledge across different stages; it was well-suited to exploratory studies focused on co-created tools and real-world public service innovation (Creswell, 2013; Cohen & Crabtree, 2006). This methodology provided depth to the inquiry into process nuances, stakeholder needs and context-specific implementation challenges.

The scope of the research is intentionally delimited. It does not seek to evaluate the effectiveness of the design sprint method as such, and it does not transfer findings to all policy-making contexts. Instead, it produces insight into a bounded subset of Policy Design Sprint practice: how outcomes can be evaluated in complex governmental settings. This research focuses on the post-sprint evaluation practices, not the quality of the design or facilitation of the sprint process. Geographically, the study is limited to Latvia, which reflects the author's ongoing interactions with local governmental institutions. Although some findings may be transferable to similar governance contexts, any transfer to broader settings should be undertaken cautiously.

Thematically, the study has focused on a number of key evaluation themes. These include societal impact (the extent to which outcomes align with public values such as social equity and pursuit of the common good), democratic legitimacy (the extent of participation and fairness), systemic coherence (the extent to which the outcomes align with existing policies and systems in place), behavioural effectiveness (the likelihood of adoption and public reaction from the implementers), and future readiness (the extent to which the outcome will be resilient in future scenarios). The development of the tool is informed by results of the research (i.e. responses to the survey and findings of semi-structured interviews as well as input from the expert facilitators engaged for the prototype walkthrough); the initial outcome was validated with the help of experts able to provide their insights from the experience of facilitating Policy Design Sprints.

The research literature review was intentionally interdisciplinary. It draws on Service Design (particularly public Service Design and co-creation processes) literature (Stickdorn et al., 2018), Innovation Studies (particularly focus on the diffusion of innovation in

bureaucracies) literature (Mulgan & Albury, 2003), Legal Design (specific on clarity of policy and institutional malleability) literature (Haapio & Hagan, 2021), Systems and Futures Thinking (particularly in relation to long-term resilience and sustainability of systems) (Edquist, 2011; OECD, 2020; Government Office for Science, 2024), and behavioural Economics (specifically on the importance of integrating behavioural insights into policy implementation) (Thaler & Sunstein, 2008). These perspectives informed the selection of what evaluation criteria might be and the design of the contours of the framework.

The empirical scope is intentionally narrow to ensure depth of analysis: the three-phased research methodology is designed to yield rich contextual data, not a broad statistical generalisation. Some identified data gaps (due to confidentiality or organisational reasons) do not detract from the development or validation of the tool. The aim was not to produce a statistically generalisable document but useful and transferable insights and a validated tool that can be modified by participants to study possible future participatory policy implementation.

In sum, the scope of the research is broad in theoretical coverage and focused on practical application. The research concentrates on the development and empirical testing of a working evaluative tool for evaluating the outcome of Policy Design Sprints. By grounding the framework in interdisciplinary theory and validating it against real-world implementations, the study ensures that the contributions are both academically substantive and practically relevant.

1.5 Thesis structure overview

The thesis is organised into six main chapters (excluding Conclusion, List of References and Appendices), as follows:

Chapter 1: Introduction. Sets the research context and foundations. It outlines the background of Policy Design Sprints and related design and innovation fields, presents the research motivation, objectives, and questions, and defines the scope and limitations of the study. This chapter also includes ethical considerations and an overview of the thesis structure.

Chapter 2: Literature Review. Provides an extensive synthesis of existing scholarship on Service Design, Policy Design, Innovation, Legal Design, Systems Thinking, Futures

Thinking, behavioural Economics, and Agile Governance. Each sub-section corresponds to a conceptual dimension of the evaluation framework developed later in the thesis. The review builds the theoretical foundation for the research and identifies key gaps in knowledge that the study addresses.

Chapter 3: Research Strategy. Describes the research approach and methods. It explains the qualitative case-study design, justifying its suitability for the exploratory research questions. The chapter details data collection techniques (including survey, semi-structured interviews, expert walkthroughs and document analysis), participant selection, and the rationale behind the selection of techniques as well as participants. It also explains the iterative process used to develop and refine the evaluation tool.

Chapter 4: Results and Findings. Reports the results of applying the framework in the two real-world Policy Design Sprints. It discusses how the framework was used in the post-sprint evaluation phase and how sprint participants and decision-makers responded to it. The chapter analyses the empirical findings, identifying strengths and limitations of the framework in practice. It also outlines policy recommendations and suggestions for integrating structured evaluation mechanisms into broader public sector innovation strategies.

Chapter 5: Design Works. By grounding design decisions in empirical data and interdisciplinary theory, this chapter presents the iterative design process of the Better Outcome evaluation tool. The chapter chronicles how initial ideas from theory and earlier findings were translated into a concrete assessment instrument through successive refinements. It describes the tool's structure and explains the theoretical rationale for each evaluation dimension. The visualization of results and considerations of usability and ethics in design are also discussed.

Chapter 6: Discussion and Implications. Chapter interprets the study's findings in light of the literature and research objectives. It explains the implications of the central results — notably the persistent implementation gap — and explores how integrating systems, futures, and behavioural perspectives helps to address this challenge. The chapter highlights both the theoretical and practical contributions of the work, including guidance on how to apply the "Better Outcome" tool in practice. It also touches on directions for future research.

In terms of the overall structure of the thesis encompasses the interdisciplinary and applied nature of the research by offering a straightforward narrative that leads to the conclusion of a tested framework to evaluate the outcomes of Policy Design Sprints.

1.6 Ethical consideration

The research engaged a range of participants, including representatives from governmental institutions, sprint facilitators, and public sector participants. All participants were informed about the purpose and essence of the study, as well as its voluntary nature. Consent documents were signed in advance, outlining participant rights, confidentiality, and data handling, and participants could withdraw from the study at any time without any negative implications.

Privacy and anonymity were consistently emphasised. Names and identifiers were removed from (or not used) in any written record, interview transcripts, and summaries. Roles, departments, or any other detail that would identify the participant were not provided to prevent any privacy concerns. Digital recordings, documents, and any files accessed during the research were stored on a password-protected computer, with only the researcher having access to it.

Participation was completely voluntary, and there was no coercive or incentivised element to the study. There was no compensation or other incentives offered; the sole consideration was to create a respectful, safe space for constructive feedback. Throughout the evaluation process, the researcher remained neutral, refraining from expressing personal beliefs or ideological agendas. Instead, the focus was on generally accepted public policy values – such as fairness, civic engagement, practicality, and long-term sustainability – relevant to all stakeholders.

The nature and context of government was considered. When the research or feedback was shared with participants at the institutions involved, the feedback was shared anonymised and framed using language that was constructive and able to support future growth as opposed to being negative and critisising the past performances. The researcher adhered to the ethics of research and scholarly work throughout the entire project, including areas accepting informed consent, confidentiality, and neutrality. This commitment to ethics helped ensure both participant well-being and the integrity of the research findings.

2. LITERATURE REVIEW

2.1. Design Sprints in Public Policy: Promise and Challenges

In recent years, design-driven approaches such as design sprints have permeated public sector innovation labs and policy design initiatives around the world. Adapted from the tech and product development realm (e.g. the Google Ventures sprint model), these short, intensive co-creation processes condense research, ideation, prototyping, and testing into just a few days. Governments have embraced design sprints as a way to generate creative solutions to "wicked" public problems, increase stakeholder engagement, and break down bureaucratic silos in policymaking (Bason, 2018; Kimbell, 2015a). A typical Policy Design Sprint brings together civil servants, experts, and citizens to collaboratively develop a prototype solution, such as a new service blueprint, a policy mock-up, or a user journey map—to address a pressing issue (Mintrom & Luetjens, 2016a; Gustetic et al., 2020). The appeal lies in sprint workshops providing a safe space for experimentation away from day-to-day constraints, allowing participants to think outside conventional routines (Bason, 2018; Brown, 2009).

Policy analysts see policymaking as a cyclical process with phases of agenda-setting, formulation, implementation, and evaluation (Howlett et al., 2009). A Policy Design Sprint can be applied in all of these phases. In the agenda-setting phase, it brings stakeholders together to identify key issues. In the formulation phase, a sprint provides the ability to quickly prototype and test assumptions. In the implementation phase, it can assist in developing and testing service delivery models. Finally, in the evaluation phase, the sprint can establish a space to reflect, allowing for early changes to be made. This allows for a learning loop that makes decision making more timely and responsive (Shiffman, 2008). Therefore, we can qualitatively assess the effectiveness of a Policy Design Sprint to link the stages of the policy cycle.

Design sprints promise not only speed but also tangible outputs that can communicate ideas in ways traditional policy analyses cannot.

Firstly, the tangible design artefacts produced (prototypes, storyboards, visual scenarios) help improve shared understanding across diverse stakeholders. A novel policy concept represented as a prototype or storyboard is more concrete and credible to decision-makers

than abstract text, easing communication up and down hierarchies (Bason, 2018). For example, a visual journey map of a citizen experience can create a shared vision of possible future service and is likely important for commensurate buy-in across departments (Kimbell, 2015a).

Secondly, prototypes create opportunities for experiential learning and early feedback. Unlike purely analytic approaches that remain theoretical until implementation, a design sprint's rapid prototypes can be tested (even if only in simulations or role-play) to gather immediate input from users or experts (Dow et al., 2012). This hands-on testing, even if informal, lets teams identify flaws or improvements before significant resources are committed.

Thirdly, the sprint format can enhance stakeholder engagement and democratic participation by making policy ideas more accessible. A mock-up or interactive model is often easier for non-experts (citizens, frontline staff) to grasp and respond to than a dense policy brief. The UK Policy Lab, for instance, found that using simple prototypes (like simplified forms and storyboards) helped prompt more open citizen feedback on how a policy might affect them, whereas abstractconsultation questions had yielded limited input (UK Policy Lab, 2019). This is one way in which sprint artifacts enable non-expert, lay stakeholders to engage in policy design while potentially increasing legitimacy of outcomes. Also, by testing ideas at a small scale, sprints can diminish risk associated with innovation: weak ideas are revealed before they are made formal; and strong ideas can build evidence and support (Pisano, 2020).

However, despite these promises, a critical gap persists: How do we know if a design sprint's solution is actuallybetter than the status quo or an alternative policy approach? The very features that make sprints attractive—speed, informality, and focus on prototypes—also make rigorous **outcome evaluation** difficult. Many government innovation teams struggle to track what happens after the sprint workshop ends (Kimbell & Bailey, 2017). There is often an "evaluation vacuum" once participants return to their regular jobs (Bason, 2010), meaning it may be unclear whether the prototype solutions ever get implemented or lead to measurable improvements in society. Sprints in Product or Service Design typically conclude with immediate user testing of the prototype and iterative refinement (Knapp et al., 2016), but in the policy context, such near-instant validation is not always feasible for a variety of reasons, like security, type of intervention,

etc. Policy interventions often require legislative changes, budget approvals, or long-term roll-out, so a prototype policy cannot be fully *implemented* and user-tested within the sprint timeframe is often problematic. Furthermore, policy outcomes are diffuse and affect broad populations and systems, not just individual users in a controlled setting. A design sprint might produce a brilliant idea on paper, but whether that idea translates into real-world impact is a complex question that may not be answered until months or years later (Sanderson, 2002). As a result, public sector sprints risk being declared "successful" based on anecdotal creativity or participant enthusiasm alone, rather than evidence of tangible public value (Lewis et al., 2020; Olejniczak et al., 2020). This gap between the **process outputs**(exciting prototypes, engaged stakeholders) and **outcome evaluation** (real-world change) is increasingly noted in the literature (Clarke & Craft, 2019; McGann et al., 2021). Overall, while design sprints in government show great promise, the **challenge of evaluating their outcomes** remains largely unmet. This literature review addresses that gap by examining how one might define a "better" policy outcome in sprint contexts and what frameworks or criteria can be used to **benchmark** a sprint's results against the current situation or alternative solutions.

2.2. Defining Criteria for a "Better" Policy Outcome

To determine whether a Policy Design Sprint actually yields a *better outcome* than the status quo, one must first clarify what "better" means in context. In public policy, outcomes can rarely be reduced to a single metric; they encompass multiple dimensions of value and performance (Moore, 1995; Newcomer et al., 2015). A review of relevant literature suggests several key criteria for evaluating the success of a sprint-generated policy innovation: **public value and legitimacy**, **stakeholder engagement**, **feasibility and alignment**, and **user-centric quality**, among others. These criteria move beyond simple output measures (like number of ideas generated) and toward assessing the sprint solution's value in the real world. Crucially, they reflect the idea that a "better" outcome should contribute to societal goals and be implementable in practice, not just novel or cost-effective on paper.

Public value and societal impact. Mark Moore's (1995) concept of *public value* provides a foundational lens for outcome evaluation in the public sector. A "better" policy outcome should create public value – that is, produce benefits or improvements valued by society

and increase citizens' trust in government. Design sprints are often justified as tools to co-create public value by focusing on user needs and creative solutions (Bason, 2010; Ansell & Torfing, 2014b). Thus, a basic criterion for success is whether the sprint's proposal stands to improve societal outcomes compared to the current policy. For example, does the new solution promise greater effectiveness in solving the targeted problem (e.g. higher uptake of a public service, reduced processing times, improved health indicators)? Does it address an unmet need or correct a failing of the status quo (Borins, 2014)? In short, **effectiveness and impact** are core: a sprint outcome must demonstrate potential to achieve the policy goals it was created for, better than existing measures do. Sometimes this involves second-order effects – for instance, a policy innovation might aim not only to deliver a service more efficiently but also to increase citizen satisfaction or trust in government processes (Bryson et al., 2014). These broader impacts can be harder to measure immediately, but they are central to judging public value. A design sprint outcome that looks efficient yet undermines democratic values or public trust would not truly be a better outcome (Weber & Rohracher, 2012). As Weber and Rohracher (2012) note, in the context of innovation policy, "better" must be legitimised in the eyes of stakeholders and society, not just calculated in cost-benefit terms. In evaluating sprint results, this means considering qualitative value – *are citizens better off, and is governance strengthened?* – not merely quantitative targets.

Democratic legitimacy and stakeholder acceptance. A closely related criterion is the legitimacy of the proposed solution and the degree of stakeholder support behind it. Because public policies ultimately need political and public buy-in to be viable, a sprint outcome that lacks legitimacy cannot be deemed a success, no matter how innovative. Legitimacy here refers to both process and outcome legitimacy (Clarke & Craft, 2019; Lewis et al., 2020). Policy Design Sprints, by involving cross-functional teams and sometimes citizens, can enhance process legitimacy through co-creation(Blomkamp, 2018). But outcome legitimacy requires that the solution aligns with community values, ethical norms, and has support from key stakeholders (electorate, interest groups, implementing agencies). **Benchmarking legitimacy** involves asking whether the new idea would be considered fair, equitable, and in the public interest by an informed public (Weber & Rohracher, 2012). For instance, if a sprint develops a digital service to replace an existing in-person process, one must evaluate if this change is inclusive and acceptable:

Will it disadvantage any group (e.g. those with low digital literacy) or invite political opposition? If a prototype policy is very efficient but erodes transparency or excludes certainstakeholders, it may score poorly on legitimacy. In other words, *"better" means better for democracy, not just better for efficiency*. Ensuring this requires evaluating power dynamics and equity implications of the sprint outcome (Lewis et al., 2020). Scholars have argued that public sector innovation should be assessed not only on its technical merits but on how it supports democratic values (Ansell & Torfing, 2014b; Olejniczak et al., 2020). Thus, a thorough evaluation of a sprint concept would consider stakeholder perspectives: Was it co-created with the people it affects (stakeholder engagement)? Do those who must implement or comply with it find it acceptable? The participation of stakeholders in developing the solution can itself be an indicator of likely legitimacy and smoother implementation (Tõnurist et al., 2017). A policy sprint outcome emerging from a collaborative, user-centered process is expected to carry more legitimacy and stakeholder"energy" than one designed top-down (van Buuren et al., 2020b). These factors should be explicitly assessed as criteria in judging the outcome.

Feasibility and strategic alignment. Another critical dimension is whether the sprint-generated solution is **practicable** and aligns with the broader policy context and goals. Public policy ideas, no matter how creative, only deliver value if they can be implemented in the real world (Howlett, 2023; Peters, 2018a). Therefore, should ask: *Is the proposed solution technically, administratively, and politically feasible to carry out?* This requires thinking through the resources and capacity, legal and regulatory changes, and potential obstacles. For example, if a sprint proposes a digital tool, does the institution have both technological infrastructure and skills to launch it? If the idea calls for coordination between different institutions, is there an existing mandate or mechanism to support that? The literature on policy implementation stresses that many innovations fail due to institutional or operational hurdles rather than flaws in the idea itself (Dunn, 2017; Klein Woolthuis et al., 2005). A sprint outcome might look promising in a workshop setting but might encounter "system failures" (Klein Woolthuis et al., 2005) such as lack of funding, legislative gridlock, or organisational inertia. As part of outcome evaluation, one should measure these implementation risks and consider whether the sprint's product has been designed with them in mind.

Closely tied to feasibility is **strategic alignment**: how well does the new solution align with the overarching policy objectives and the environment? Innovation literature notes that pilots and prototypes can sometimes be orphaned if they do not fit the strategic priorities of the host organisation or government (Clarke & Craft, 2019). Thus, “better” means not only novel but also relevant and aligned. If a sprint outcome addresses a problem deemed important by stakeholders and fits within current political priorities or reform agendas, it stands a much better chance of adoption (McGann et al., 2021). Conversely, a brilliant idea misaligned with the institution's mandate or conflicting with other policies may never see the light of day. One way to evaluate alignment is to see if the sprint explicitly tied its goals to high-level policy goals or pain points identified by leadership. Research on policy design suggests that successful innovations often have champions and a clear link to strategic objectives (Mintrom & Luetjens, 2016b; Nogueira & Schmidt, 2022). In evaluating the sprint outcome, criteria might include: Does it complement (or even more importantly, does not contradict) existing policies and programs? If the answer is yes, the outcome is stronger on this dimension of "better." In summary, practicality and alignment ensure that the prototype is not just ideologically desirable but *viable*: it can realistically be executed and sustained within the system it’s meant for.

User-centric design and experience quality. Since design sprints draw from human-centered design, another important criterion is the quality of the solution from the **end-user’s perspective**. Even a policy that is effective on paper can fail if citizens or front-line workers find it too difficult to understand, use, or accept (Norman, 2013; Blomkamp, 2018). Service Design literature emphasises evaluating both the functional utility of a solution and the user experience it creates (Parker & Heapy, 2006). For sprint outcomes, this means checking: Is the proposed policy or service user-friendly and accessible? Does it simplify and improve the user journey compared to the current state? For example, if the sprint’s output is a redesigned permit application process, one would evaluate if the new process is easier and faster for citizens to complete,and whether it reduces pain points present in the old process (Chan et al., 2025). A *better outcome* should score higher on user satisfaction and accessibility. Additionally, human-centered policies often aim for **empathy** – being responsive to user needs and contexts (Brown, 2009). An evaluation can examine how well the sprint team understood and incorporated actual user

insights: Did they do user research? Does the solution address real user pain points or just presumed needs? A policy innovation built on solid user research is likely more relevant and thus "better" from the standpoint of those it serves (Blomkamp, 2018). In absence of the ability to pilot test the solution broadly, proxies like expert reviews or small-scale simulations with users can be used to measure expected user experience (Nielsen, 1994). Ultimately, a design sprint outcome that is efficient, aligned, and potentially impactful could still fail if it is not **usable** or acceptable to the users it has been designed for – making this an indispensable evaluation criterion. As Bason (2018) argues, successful policy design marries the "functional" and the "human": the litmus test of better policy is often whether people actuallyunderstand it and find value in it.

In practice, these criteria (public value, legitimacy, feasibility, alignment, user experience) provide a multi-dimensional profile of a Policy Design Sprint outcome. Each represents a question that any claim of a "better" solution should answer: *Better in what ways, and for whom?* Traditional evaluations in government tended to focus on efficiency or cost-benefit alone, but as modern scholars note, complex innovations demand a broader lens (Sanderson, 2002; Patton, 2011). As anexample, a narrowly framed metric, such as cost savings, may be missed if the solution undermines equity inherent in previous service delivery or if a potential solution simply lacks support by the political class. Formulating a framework whereby success can be defined across several dimensions allows policymakers to make a more holistic judgement about the outcomes of a sprint. Recent public administration literature speaks to this need for composite success measures – e.g., van Buuren et al. (2020b) emphasise legitimacy and stakeholder ownership as key outcomes of co-design, not just the technical solution generated. In context of this thesis, a **balanced criteria framework** allows for a structured comparison of a sprint's outcome against the status quo or another solution which essentially is a multi-criteria benchmark exercise. Before turning to benchmarking, it is worth noting that not all criteria will weigh equally in every case; part of the evaluator's task is to prioritise what "better" means for the specific policy domain (Howlett, 2023). For instance, in a highly sensitive policy area, legitimacy might trump efficiency, whereas in a service delivery context, user experience and efficiency might be crucial. Having explicit criteria in these categories ensures that evaluation moves beyond intuition and includes the aspects of public value that matter.

2.3. Benchmarking Sprint Outcomes for Accountability and Learning

Benchmarking in this context refers to comparing the proposed sprint outcome directly against a baseline (the current situation) or against an alternative policy solution, using the criteria defined above. The aim is to determine whether the sprint's innovative solution truly constitutes an improvement – in other words, to substantiate the claim that it is a "better outcome." Benchmarking provides a structured way to answer the question: *Would this new idea perform better than what we have now?* This approach is crucial for both **accountability** (justifying the value of design interventions in government) and **learning** (understanding what works and why in policy innovation). Unlike private-sector product sprints where market feedback provides a clear benchmark (e.g. increased user engagement or sales), Policy Design Sprints operate in a realm where feedback is diffused and success can be subjective. Hence, deliberately setting up comparisons can inject some rigour into the evaluation of policy prototypes (Mercure et al., 2021a).

A first step in benchmarking a sprint outcome is to establish a **baseline measurement** for the current policy or situation. For example, if the sprint addressed slow processing times in a permit system, what is the current average processing time and user satisfaction level? These baseline metrics form the yardstick against which the new design is measured (Haynes et al., 2012). Ideally, these metrics are identified *before* or during the sprint, so that the team is clear on what "success" should look like (Olejniczak et al., 2020b). This echoes best practices in program evaluation: one must define indicators and gather baseline data to enable a before-and-after or us-vs-them comparison (Newcomer et al., 2015). In real-world sprint practice, however, teams do not always have the luxury of robust baseline data. Still, even approximate figures or qualitative baselines (e.g. "citizen complaints are frequent about X process") are better than none. **Without predefined criteria and baseline measures, any claim that the new design is "better" remains subjective**. Thus, one recommendation from both the literature and practice is that policy sprints should begin by clarifying the problem and how current policy performance is judged (e.g. what are the pain points and their magnitude), which then enablesbenchmarking improvements (Huić et al., 2023).

The benchmarking process includes comparing the *expected* performance of the sprint's proposed solution to the baseline or *status quo*. This often requires some degree of projection, since the sprint outcome is typically a prototype not yet something that is implemented and measurable. Tools from policy analysis can assist here: for instance, if the criterion is efficiency, the team can estimate potential time or cost savings from the new process compared to current data. If the criterion is compliance or uptake, one might use analogies from similar policies or expert judgment to predict whether the redesign would improve those rates. For example, if the proposed new permit process is online and user-friendly, one might expect higher completion rates; a benchmark could be "increase permit applications completed within one week from 60% (current) to 80% with the new design." Sometimes, quick **experiments or simulations** can be conducted to generate comparative data. For instance, a sprint team could pilot their new form with a small group of users and compare completion times to the old form (Radnor et al., 2012). Such quick tests, even if not fully scientific, provide evidence to benchmark outcomes. This approach aligns with the idea of *rapid evaluation*, where rather than waiting for long-term results, teams gather indicative data in a short cycle (Rowe, 2019). Andy Rowe's Rapid Impact Evaluation method, for example, engages stakeholders and experts to estimate the impacts of innovation in a workshop format, yielding comparative judgments much faster than traditional evaluations (Rowe, 2019). A design sprint could incorporate a mini"impact evaluation session" at its conclusion: key stakeholders review the prototype and score it on the success criteria relative to the status quo. This kind of structured debrief can be a form of benchmarking, capturing whether knowledgeable observers believe the idea outperforms current practice (and why or why not). While not as rigorous as a field experiment, it grounds the assessment in systematic comparison rather than hype.

Academic work on policy innovation evaluation supports this push for comparative frameworks. Mercure et al. (2021a) propose a **"risk-opportunity analysis"** for transformative policy design that generalises cost-benefit analysis to account for deep uncertainty and dynamic change. Their approach essentially benchmarks new policy proposals against a reference scenario (business-as-usual) across multiple dimensions, including not just economic efficiency but also risks (downside scenarios) and opportunities (upside scenarios). This resonates with the multi-criteria outlook of the sprint evaluation: instead of a single metric, one examines a range of outcomes and their

distributions. Applied to a design sprint, a risk-opportunity lens would encourage teams to consider: In what ways might the new policy outperform currentpolicy, and in what ways might it underperform or introduce new risks? Such nuanced benchmarking is valuable because an innovation might excel on some criteria (say, equity and satisfaction) while initially lagging on others (perhaps cost or time to implement). Rather than declaring the sprint outcome simply "better" or "worse," evaluators can identify **trade-offs** explicitly. For example, a new policy might make services more accessible (better equity) but also slightly more costly to deliver in the short term. Whether that constitutes a net better outcome can then be a deliberate decision, ideally informed by stakeholder values and political priorities (Peters, 2018a). Comparative evaluation thus facilitates a more transparent decision-making process about adopting sprint results, turning subjective impressions into documented comparisons.

Another reason benchmarking is important is **accountability and learning for the innovation process itself**. Policy innovation labs and design teams need to demonstrate that their methods add value (McGann et al., 2018). By comparing outcomes, they can produce evidence of impact (or lack thereof). If a design sprint's solution, when benchmarked, shows significant improvement over the old approach on key metrics, it validates the investment in the sprint and can justify scaling the solution. If not, it yields lessons for why the new idea did not surpass the old (perhaps the problem was mis-framed or the prototype was too incremental) – knowledge that can inform future projects. This reflective practice is akin to the *test-learn-adapt* cycle promoted by advocates of experimental government (Haynes et al., 2012). Indeed, governments are increasingly encouraged to use trials and A/B tests for policies, but design sprints could be a front-end to such trials by generating ideas to compare. While full **RCTs** (randomised controlled trials) or longitudinal evaluations might be impractical immediately post-sprint, the ethos of experimentation can be maintained through simpler benchmarking exercises that still foster learning (Patton, 2011). In some cases, if time and scale allow, a sprint-developed solution could be implemented in a pilot region while another region continues with the status quo, creating a natural experiment. Short of that, even **simulated comparisons** and expert elicitation (as in Rapid Impact Evaluation) can create a feedback loop. The key is that *benchmarking makes the success criteria explicit and observable*, rather than relying on the excitement of the sprint event as a proxy for success. This is important for the

legitimacy of design-led innovation in the public sector: political and public supporters of an innovation lab will expect real-world results, and comparative evidence is compelling in that regard (OECD, 2019).

It is also worth noting that benchmarking a sprint outcome need not only compare against the existing status quo; it can also compare against alternate solutions or previous attempts to solve the same problem. For example, if a ministry was considering two competing proposals (perhaps from different teams or from earlier work), then it could also benchmark the sprint proposition against its alternative by comparing costs, acceptability and impact on a range of indicators. Such benchmarking situates the outcome of a sprint in relation to another option and may shed light on the relative strengths and weaknesses of this route of action.

Such analysis benefits from frameworks like **realistic evaluation** (Pawson & Tilley, 1997) which emphasise context: understanding *what works for whom under what conditions*. A sprint outcome might be better than status quo in the current context, but perhaps an entirely different policy approach could achieve even more – evaluators should remain open to that possibility. The literature on evidence-based policymaking encourages considering multiple options and using criteria-based comparison to choose among them (Vedung, 1997; Dunn, 2017). In conclusion, benchmarking is a flexible but fundamental activity in the sense that in comparing the proposed policy change from the sprint either to the status quo or against another option in terms of pre-determined criteria, policymakers can make a rational argument about whether this sprint has given them a better outcome. This approach instills a discipline of **evidence and comparison** into the creative chaos of design sprints, helping bridge the gap between innovation and evaluation.

2.4. Integrating Systems, behavoural, and Futures Perspectives in Evaluation

Evaluating Policy Design Sprint outcomes requires not only criteria to build comparisons, but also a broad perspective that situates the innovation within **larger systems, human behaviours, and future contexts**. Traditional evaluation criteria (like those outlined above) can be enriched by insights from policy process theories and design-related disciplines that emphasise different lenses: Systems Thinking, behavoural science, and futures/foresight. By applying these lenses, evaluators can identify potential pitfalls or

strengths of a sprint outcome that might be missed if we only look at the immediate metrics. In effect, these perspectives ensure that a sprint's "better outcome" is resilient and robust in the **real-world ecosystem** where it will live. Recent research and practice in policy design highlight the importance of such multidisciplinary integration (Romme & Meijer, 2020; Olejniczak et al., 2020). This research explores each of these perspectives and their relevance to sprint outcome evaluation.

Systems Thinking and policy context. Policies operate within complex systems of institutions, regulations, and stakeholders. A design sprint outcome, typically a prototype solution, will have to contend with this complexity upon implementation. Systems Thinking urges us to consider the broader context: what are the enabling or inhibiting factors in the system that will affect the success of the new policy design? One useful framework from innovation policy is the **system failure framework** (Klein Woolthuis et al., 2005), which categorises different kinds of failures (market failures, institutional failures, network failures, etc.) that can hinder an innovation's implementation. Applying this to a sprint outcome, evaluators might examine: Does the proposed solution address known systemic failures or does it risk running into them? For example, suppose a sprint's idea requitutires high cross-institutional cooperation, and previously there'sbeen an institutional failure (silos and poor coordination) in this domain. In that case, the evaluation should note that risk and perhaps judge the outcome less "ready" unless it includes a strategy to overcome that silo issue. Conversely, if the sprint outcome cleverly navigates an infrastructure limitation or simplifies a regulatory bottleneck, it earns points for tackling a system-level barrier. Using frameworks like this essentially adds a **"readiness for system change"** criterion to intended evaluation. It ensures we look beyond the artifact to the environment: Are there policy or legal changes needed to make this work? Are power dynamics in the system aligned, or will there be opposition? The **Advocacy Coalition Framework (ACF)** offers another lens here. ACF reminds us that policy change often depends on shifts in stakeholder coalitions, interest group support, and prevailing belief systems (Pierce et al., 2020a). If a sprint outcome implies significant policy change, the evaluation should assess whether it has garnered support (or even more importantly, not denial) from key stakeholders and whether it aligns with (or challenges) prevailing policy paradigms. For instance, a very radical proposal might be "better" in outcome terms but completely at odds with the current government's ideology, making its

actual adoption unlikely (Howlett, 2020a). An evaluator informed by ACF would note the **political feasibility**: does the idea fit the current policy subsystem's values? Does it have champions in the advocacy coalition, or could it form a new coalition? These system-oriented questions complement the earlier feasibility and alignment criteria by explicitly examining the *dynamics of change* in the policy system. In practice, a thorough sprint evaluation might include a brief stakeholder analysis and context scan – effectively a systems mapping – to identify external factors that could affect the outcome's success (Williams & Hummelbrunner, 2010; Jones & Bowes, 2017). This helps avoid the trap of evaluating the prototype in a vacuum. By looking at system readiness, evaluators ensure that a "better" outcome on paper truly has a pathway to be better in reality.

Behavioural insights and assumptions. Every policy design carries implicit or explicit assumptions about how people will behave. For example, a new online service assumes that citizens will choose to use it, that they have internet access and trust the digital platform. A public health policy might assume people will respond to information campaigns or incentives in certain ways. Behavioural science has taught policymakers that such assumptions often fail if not empirically validated (Datta & Mullainathan, 2014; Thaler & Sunstein, 2008). Design sprints, being human-centred, are usually mindful of user needs, but time constraints may prevent deep validation of behaviour change aspects. Therefore, when evaluating a sprint outcome, it is valuable to apply a **behavioural lens**: ask what assumptions about human behavour or organisational behaviour underlie the solution, and are these plausible? Olejniczak et al. (2020b) provide a framework for comparing the behavioural assumptions of policy tools, essentially encouraging policy designers to spell out how they expect their intervention will change behavour and to check if those assumptions align with evidence or theory. For instance, a sprint outcome might assume that providing more information will lead citizens to make a desired choice – a classic assumption that behavioural economics often finds overly optimistic unless combined with nudges or simplification (Knetsch, 2011; Shafir, 2013). An evaluator (sprint team or sprint facilitator) could flag this: *willinformation alone suffice, or are additional behavioural nudges needed?* Maybe the new process should incorporate a default option or social incentive to truly achieve the desired behavour change (Thaler & Sunstein, 2008). If the sprint outcome neglects such considerations, its effectiveness could be in question. On the other hand, if the design explicitly uses behavioural insights (say, it simplifies a form to

reduce friction, or uses timely prompts), that's a positive indicator that it will work as intended. In criteria framework, this falls under effectiveness, but the behavioural lens gives a more fine check on *mechanisms of change*. Importantly, evaluating behaviour assumptions also covers internal behaviours – for example, if implementation relies on public servants to adopt a new tool, does the design account for their incentives and routines? Traditional evaluations might not consider that level of detail, but a design-oriented evaluation would, because it recognises policy designs must fit human behaviour to succeed (Olejniczak et al., 2020b). By incorporating behavioural science, we guard against the risk that a "brilliant" solution fails simply because people didn't behave as expected. To illustrate, imagine a sprint produces a policy that relies on citizens opting in to a new program. If historically opt-in rates are low, an evaluator might suggest the outcome isn't truly better unless changed to opt-out (leveraging the default effect). In summary, the behavioural perspective asks: *Is the sprint outcome aligned with how real people are likely to act?* If yes, it strengthens confidence in the outcome; if not, the evaluation might recommend modifications or further testing.

Futures and foresight thinking. Policy design should also consider the future context into which a solution will be introduced. A design sprint often focuses on solving today's problem, but by the time a policy is implemented, conditions may have evolved. **Futures Thinking** encourages designers and evaluators to explore how robust an innovation is under various future scenarios (Government Office for Science, 2024). In evaluating a sprint outcome, a futures lens would pose questions like: Will this solution still be effective 5 or 10 years from now? What external changes (technological, social, economic) could occur that either bolster or undermine its impact? For example, if a sprint in 2019 designed a policy around encouraging carpooling, an unforeseen pandemic in 2020 would drastically alter the context of that policy. While we "cannot predict the future exactly," we can consider plausible scenarios (Schwartz, 1996; Wilkinson & Kupers, 2013). Scenario analysis as part of evaluation could reveal, say, that the proposed solution is highly sensitive to economic growth (it works well if budgets grow, fails if they shrink) or that it assumes a stable political environment. This doesn'tnecessarily disqualify the outcome but provides nuance: perhaps the idea is a great short-term fix but not a long-term one. Or maybe it's adaptable to multiple futures, which would be a strong point. This suggests that sometimes it makes more sense to discuss the adaptability to the unknown rather than

readiness to trending future aspects. *Anticipatory governance*literature suggests that policies should be designed with flexibility to adjust as conditions change (OECD Observatory of Public Sector Innovation, 2023). An evaluation step might check if the sprint outcome has such flexibility or if it's a one-shot solution that could become obsolete. In concrete terms, Futures Thinking in evaluation might involve reviewing the sprint idea against megatrends or uncertainties (e.g., ageing population, climate change, evolving technology) to see if it remains relevant or needs additional safeguards (Dufva, 2019; Cambridge Assessment, 2023). If a futures toolkit (Government Office for Science, 2024) had been used in the sprint, that would likely improve the design's resilience; if not, the evaluator might simulate a quick foresight exercise as part of the review. Considering futures also ties back to sustainability: a "better outcome" should ideally be sustainable over time, not just a quick win. Does the policy innovation have a funding model that can last? Is it scalable if conditions demand? These are forward-looking criteria that complement the present-focused ones. By integrating a futures perspective, we ensure that the evaluation is not myopic; it accounts for the fact that policies often need to function in a changing world.

Bringing these perspectives together, we arrive at a **multi-dimensional evaluation approach** that is well-suited to the complexities of public sector design sprints. Systems Thinking contributes an understanding of context and structural factors, behavioural insight checks the human factor and realistic mechanics of change, and Futures Thinking tests robustness over time. Adding to the earlier criteria and benchmarking, this is a complete framework within which to assess sprint outcomes. For instance, let's say a sprint outcome ranks highly on immediate criteria (it's effective, it'sfeasible, it's user friendly, etc.). A systems check could still identify a risk of non-compliance with existing regulations - leading to a recommendation to change the legal framework or the design. A behavoural check might lead to a recommendation for default enrollment in order to increase uptake. A futures check could verify that the solution is flexible enough to work in different contexts and settings, which would increase confidence in the outcome overall. In contrast, without these lenses, one might prematurely celebrate the outcome and face surprises down the road.

Notably, the need for such integrated evaluation is being recognised in both scholarship and practice. Romme and Meijer (2020) argue that **design science should be applied in**

public policy, combining design creativity with scientific evaluation's rigour in an iterative cycle. They and others envision policy design as an ongoing process of hypothesis (design idea), experiment (implementation/pilot), and evaluation (learning), much like in design thinking but with more emphasis on evidence. Policy labs – dedicated innovation teams in governments – are increasingly exploring how to evaluate their interventions in real-time (Olejniczak et al., 2020b). Some labs are developing bespoke evaluation toolkits that bring in systems mapping, behavioural checkpoints, and rapid feedback loops to assess innovative projects (OECD, 2019; Nogueira & Schmidt, 2022). This is essentially operationalising the integration discussed. The literature review by Hermus et al. (2020) finds that while design and policy have historically been separate spheres, there is a convergence where design approaches are used for policy **formulation** and traditional evaluation methods are beingadapted to shorter, design-driven cycles. The process articulated in this review - identifying explicit criteria, benchmarking against a baseline, and using a range of lenses - helps to support that convergence by allowing a systematic way to evaluate sprint outcomes rigorously while allowing for innovation to flourish.

Overall, evaluating the outcomes of Policy Design Sprints is a multi-dimensional challenge that demands a nuancedapproach. By comparing sprint outputs to the status quo or alternatives (benchmarking) and by examining them through criteria of public value, legitimacy, feasibility, alignment, and user experience, we get a grounded assessment of whether an idea is truly "better." By further integrating systems, behavioural, and futures perspectives, we ensure that the assessment captures the complexity of real-world implementation and longevity. This comprehensive literature-informed approach turns the design sprint's mantra of "move fast and prototype things" into "move fast, but also measure and reflect on things." Ultimately, the goal is to create a feedback loop where design sprints not only generate innovative policy ideas but also build knowledge on what innovations actually deliver public value. In doing so, public sector innovators can be more accountable for results and more effective in scaling solutions that work. The literature suggests that such an approach—marrying the creative power of design with the analytical rigour of evaluation—can significantly enhance public service innovation (Bason, 2018; Patton, 2011; McGann et al., 2021). It transforms the design sprint from a one-off idea generator into an integral part of evidence-based policy development, helping ensure that the sprint'soutcomes truly make a positive difference when translated into policy action.

3. RESEARCH DESIGN

3.1 Research Strategy

This study utilised a qualitative exploratory approach based on an interpretivist-constructivist position. In this perspective, as Crotty (1998) describes, "Interpretivism" sees reality as socially constructed out of the sensemaking activities engaged in by human actors. In line with Creswell (2013, 2014), the research did not seek to confirm an effect or test a hypothesis but sought to understand better how the participants made meaning of their experiences. To this end, this notion supports the ability of participants to construct various subjective interpretations of lived social phenomena in an open-ended emergent research design (Creswell, 2013). In this interpretivist way, the researcher aimed to elicit rich, contextualisedaccounts from the stakeholders in order to understand how they made meaning of "better outcomes" of design sprints (Denzin & Lincoln, 2018).

Recognising that different participants may hold varying criteria for success, the study embraced the interpretivistrecognition of multiple, co-existing realities. For instance, one participant may view a sprint as a valuable exercise for fostering collaboration and creative thinking within the department, while another might focus on the likelihood that a prototype will be implemented (Lincoln & Guba, 1985). The researcher adopted this perspective to keep the research rooted in stakeholder viewpoints, rather than relying on external standards or rigid measurable experimentation. By allowing for interpretative flexibility, this evaluation framework can genuinely reflect how policy innovators understand their work in its various meanings.

The study also recognised a constructivist epistemology, recognising and acknowledging that, as one might expect, all knowledge is co-constructed by both the researcher and the participant (Lincoln & Guba, 1985). The researcher's ownbackground, worldviews, and engagements influenced the inquiry, necessitating a consciously collaborative and iterative research design. This approach reflected a research-through-design framework in which insights gained from each research phase informed and refined the evaluation tool in co-creative dialogue with stakeholders (Koskinen et al., 2011). Participants effectively contributed to criteria for "better outcomes," resulting in knowledge grounded in real-world cases.

The exploratory nature of this research strategy was crucial, as Policy Design Sprints typically yield intermediate outputs such as ideas, prototypes, and reframed problems—outputs that defy easy measurement via fixed indicators (Dorst, 2011; Howlett, 2020a). Traditional quantitative methods that limit the diversity of feedback, would therefore be unsuitable for capturing the details and nuances of Design Sprint processes and outcomes. Therefore, methods that aligned with theadaptable, interpretive context of policy innovation were selected.

Practically, the researcher treated the study as an exploratory investigation that gathered information from participant responses and refined interpretations in cycles (Creswell, 2013). In this sense, open prompts and broad thematic questions encouraged the interviewees and other participants of the reasearch to explain their perspectives that would inform the development of the "Better Outcome" tool. This responsiveness aligned with the calls in both design research and evaluation to be more focused on process quality and iterative learning rather than performance measurements (Patton, 2015).

In methodological terms, each research step was viewed as a complex, context-bound event, best explored inductively. This meant being highly responsive to data evolving over time instead of following a predetermined design. A flexible orientation meant that the study could capture in-depth narratives about how ideas were pursued, altered, or stalled from sprints. The survey produced breadth in perspectives, the interviews produced depth, and expert walkthroughs provided iterative feedback that shaped the emerging tool. Findings from these three qualitative methods were triangulated to account for and express the complexity of real-world policy innovation, avoiding simplistic or reductionist evaluations.

The emphasis on iterative development and flexibility as an interpretive-constructivist approach aligns with the theory of utilisation-focused evaluation (Patton, 2015). This perspective positions evaluators to continually respond to the emerging needs of the users and that the tool which is under development remains pragmatic and meaningful to stakeholders. Thus, the findings from the first phase (online survey) naturally informed the second pahse (semi structured interview questions) and both bought together ultimately informed the continuing development of the tool itself.

Overall, an interpretivist-constructivist orientation was essential to ensure that the evaluation framework was pragmatic, significant and thoughtful of stakeholders lived

experiences, success criteria, and other ideas. The multi-stakeholder and co-created spirit of this research strategy relates well with the nature of co-design in the contemporary public sector innovation (Bason & Austin, 2022; Manzini, 2015). Rather than seeking simple explanations, the underlying goal was to research for patterns and practical challenges in handling design sprint outcomes, using these insights to build a relevant and usable evaluation tool iteratively.

3.2. Research methodology

The study utilised a mixed qualitative methods design, using sequential and iterative phases. First researcher conducted an online survey to collect initial information from a broad sample of practitioners. This was followed by semi-structured interviews with key informants to explore further the findings from the survey. Lastly, expert walkthrough sessions were organised and held to assess and refine the evolving "Better Outcome" tool. By design, each phase built on the previous phase, where themes from the survey informed what questions to ask in the interviews, and the ideas from the survey and interview process impacted the development of the tool (Creswell, 2013). Exploratory, multi-method studies are recognised in qualitative research for enhancing validity and for discovering "new ways of seeing and understanding" the topic (Bernard, 1988). Cohen and Crabtree (2006) state that open-ended questions in semi-structured interviews provide the opportunity broader understanding of the topic.

Survey. The first instrument was an anonymous online questionnaire distributed to public servants who had participated in the public sector design sprints in Latvia. The call was sent via various government innovation networks and relevant email lists. Every step of participation was voluntary and confidential. The survey comprised a mix of multiple-choice and open-ended questions. The closed questions were generally factual in nature (e.g. whether any sprint ideas were implemented), while open questions invited free-text reflections on follow-up processes and challenges. The survey adopted purposive sampling: only those with Policy Design Sprint experience were targeted, thereby ensuring an "information-rich" sample. As Palinkas et al. (2015) explain, purposeful sampling involves "identifying and selecting individuals or groups that are especially knowledgeable about or experienced with a phenomenon of interest". In total, 29 completed the survey, but 21 (Participants of Policy Design Sprints) qualified for answering the questions related to this research. Given the exploratory aims, this was considered sufficient to discover

recurring patterns. (As Palinkas et al.,(2015), point out qualitative studies often rely on precedent and saturation rather than statistical power). Given the estimated number of Policy Design Sprints run in Latvia annually, and the average team size involved in each, the number of responders may be seen as relatively representable, with more than 10% of estimated individuals involved in the relevant activities taking the survey. Furthermore, the survey collected a variety of projects and thus, provided a broad "snapshot" of what post-sprint outcomes look like in practice.

Interviews. To further explore the survey results, the researcher conducted semi-structured interviews with key informants. Four interviews were completed (four of five invited) with senior figures who have led Policy Design Sprint teams. The interviewees were purposefully sampled since they had unique, first-hand experiences across multiple sprints and respective follow-ups. A semi-structured format was chosen to allow for both flexibility and guidance. The interviews adhered to a wide framework, aligned to the research questions, (e.g., outcome implementing, evaluation practices, issues), but allowed interviewees to elaborate about anything else they view as important to discuss. An example in practice here would be asking an open question such as “How to define a success of a Policy Design Sprint?”, and then follow-up questions based on their responses. The sessions were intended to be 45 - 60 minutes and recorded with consent. Afterward, the recording was transcribed verbatim. As noted by Cohen and Crabtree (2006), semi-structured interviews “provide a clear set of instructions” yet allow informants freedom to express their views, making them especially suited for exploratory fieldwork. Considering the purposeful selection of experienced sprint leaders, the interviews effectively exposed rich examples and explanations that helped interpreting the survey data.

Expert Walkthroughs. The final method involved a series of expert reviews where the emerging actual tool (the “Better Outcome” tool) was presented to three independent design sprint facilitators. The facilitators were not from the interviewed agencies but selected from a professional network of public sector innovation experts (including some with international facilitation and consulting expereince). Each expert underwent an interactive session where they reviewed the tool's criteria, definitions and use-cases. In effect, they served as peer reviewers or usability testers, providing feedback on the content and usability of the developing tool. This method is similar to expert evaluation approaches

more commonly used in design research (e.g., Rieman, 1993; Nielsen 1993), and helps to validate the fitness-for-use of an artefact. The experts were asked to 'think aloud' as they rated the potential tool by applying it to hypothetical sprint scenarios, providing feedback about clarity, completeness and usability. Verbal feedback from participants was documented through note-taking and recordings. No additional data on sprint outcomes were collected at this point, the sessions were meant to iteratively improve the tool itself. The research provided a meaningful, stakeholder-centric check of the tools design that ensured the instrument is relevant and understandable to its intended users.

Together, these three qualitative methods formed a **sequential triangulation** strategy. Each method had its ownadvantages: breadth of the survey, depth of the interview, and practical validation from the expert sessions. As noted by Shenton (2004) and Lincoln and Guba (1985), triangulation across methods and types of informants provides added credibility to qualitative findings. For instance, themes emerged in the open-ended survey responses can be confirmed or expanded in the interviews, and later explored in the expert sessions. In this way, the methods fed back to each other across an iterative cycle instead of working in isolation.

3.3. Research Context and Participants

Institutional Context. While the research was performed in the context of Latvia's public-sector innovation ecosystem, the study explored multiple examples of Policy Design Sprints within different government institutions rather than focusing on a single case study. Participating organisations included a state agency conducting a sprint on digital public services, a ministry department using sprints for legal-policy innovation, and a municipal government applying sprints to community co-design. These examples covered varied domains (digitalisation, regulatory reform, internal process innovation, etc.), allowing the study to identify cross-cutting issues rather than case-specific anomalies. In that sense, the design mirrors a multi-site field study. By sampling across institutions and policy areas, the research sought to capture common patterns and constraints of the sprint process in public settings. This broad scope was intentional: rather than generalising statistically, the goal was to understand the range of experiences of design sprint practitioners in Latvia, and to seek transferable insights for other similar innovation contexts.

Survey Participants. The online survey (Section 3.2) drew responses from 21 individuals who had taken part in at least one Policy Design Sprint. Respondents came from various roles and organisations. Because this was a focused, purposive survey, no claim is made to represent the broader population; instead, the aim was to gather "information-rich cases" of sprint experience. As Palinkas et al. (2015) note, purposeful sampling targets participants with direct knowledge of the phenomenon, maximising the richness of data. In practice, the survey was distributed through professional networks connected to government innovation programs. This sample size was judged sufficient for broad pattern-finding. Indeed, qualitative studies often rely on reaching saturation (no new themes emerging) rather than large numbers. From the 21 completed surveys, themes such as "lack of follow-through" and "need for evaluation criteria" already began to surface, indicating that the sample was capturing relevant concerns.

Interviewees. The four interview participants (coded as Interviewee 1–4) were senior project leads in their institutions, each with direct responsibility for organising or leading one or more Policy Design Sprints. They were selected because they possessed in-depth experience on the sprint process and its aftermath. Each interviewee had multiple years of public-sector service, two had overseen at least two sprints, other two have had experience in leading one sprint team. The individuals represented a mix of agencies: for example, one was a ministry official in charge of policy projects, another was a municipal leader involved in co-designing projects, etc. This diversity of institutional roles helped reveal whether themes were organisation-specific or more general. Because all interviewees were of similar high-level status, the sample was relatively homogenous in that sense. However, what varied was the *policy context* of their sprints (e.g. regulatory vs. digital services, regional vs nation wide implication). The small number (N=4) was considered adequate under a qualitative paradigm of information saturation (Palinkas et al., 2015). Guest et al. (2006) suggest that even a handful of expert informants can yield substantial insights in a focused study. In any case, each interview lasted long enough to gather detailed narratives. The researcher reviewed transcripts for recurring topics, using the interview data mainly to explain and contextualise the survey findings (rather than to generate quantitative metrics).

Expert Reviewers. In addition to organisational insiders, three **independent experts** were engaged to review the evaluation tool. These experts were experienced facilitators of policy or public-sector innovation workshops drawn from a broader network (including

international consultants and innovation lab leaders). Each expert had led numerous design sprints or similar activities in different contexts. They were not employees of the Latvian agencies studied. The intent was to bring a critical outsider's perspective. These experts reviewed the draft “Better Outcome" tool in workshop - formatonline meetings and provided open feedback. Their role parallels that of peer reviewers: they validated whether the themes and criteria derived from the Latvian cases resonated with practice in other settings. For instance, one expert noted that sprints often lack pre-defined success metrics, supporting an interview theme. Another expert suggested practical constraints (like time pressure and timing of use of the tool) that informed final edits to the tool. Including these neutral participants helped addressing potential bias or blind spots. By comparing the emerging tool against their own experiences, the experts effectively extended the findings’ applicability beyond the specific cases.

Scope and Limitations. It is important to note the boundaries of this research. No *new* design sprint was launched specifically for this study (due to time and resource limits), so there was no live pilot of the tool in action. Instead, the expert sessions served as a surrogate test. Moreover, the context remained within Latvian public-sector practice; while the tool is intended for broad use, the evidence base (survey and interviews) reflected Latvian policies and organisationalcultures. Finally, because of the qualitative approach, the emphasis was on depth and richness rather than on numerical generalizability. Wherever possible, the thesis notes overlapping themes across sources to suggest broader relevance. For example, despite different settings, all interviewees reported difficulties in sustaining momentum post-sprint. Such convergences give confidence that the findings point to common issues in policy sprints.

3.4 Data Analysis and Synthesis

All qualitative data gathered (survey open-responses, interview transcripts, expert session notes) were analysed in a systematic thematic approach. Thematic analysis is a recognised practice in qualitative research, which identifies and interprets patterns in textual data (Braun & Clarke, 2006). Analysis was conducted in two phases. First, each data source was analysed independently. Since the responses to the survey's open-ended questions and field notes (N=21 respondents) were not extensive, they were hand-coded. Simple descriptive codes were constructed to represent recurring ideas: for example, codes like *NoFollowUp* (indicating that no action was taken after a sprint) or *ChampionPresent*

(indicating that a project champion supported implementation) were used to label relevant comments. Codes were derived directly from the data (a bottom-up inductive process) and the topics of the interview (a top-down structure for consistency). The researcher used a hybrid deductive-inductive strategy; while some of the earlier codes were informed by the interview questions (e.g. *OutcomeDocumentation, PostSprintDecision*), some codes emerged from the reading (e.g. *LeadershipChange*, capturing mentions of political turnover affecting outcomes). The expert review feedback was transcribed from recordings and emerging or recurring topics were categorised thematically in a similar way (e.g. issues with *TerminologyClarity*, suggestions for *NewCriteria*).

The second phase of analysis was to integrate and synthesise across the sources. The researcher looked for core themes that cut through the survey, interview and expert input. The goal was to identify the most salient factors that influenced sprint outcomes. Several major themes arose. One common theme was one labelled *LackOfFollowThrough*; both the survey and interview data suggested that ideas from a sprint often "died on the shelf." Expert data strongly supported this theme, with experts indicating that it was a commonly known pattern. Another theme was *DivergentSuccessCriteria*; stakeholders differed on whether a sprint's success meant a concrete policy change or more intangible learning. A third theme was *StrategicAlignment*; the extent to which outputs of the sprint aligned with the institutional priorities greatly influenced implementation rate. Importantly, in some cases, the experts' feedback often stressed the practical dimension of the themes identified by insiders (e.g., time constraints on what could be evaluated). In every case, evidence occurred and was compared across multiple data sources. When the same issue arose in the survey comments, in an interview narrative, and in expert discussion, its significance was reinforced. For example, *InstitutionalFit* arose in an interview ("..it does not comply with our strategical documents") and was reiterated by experts as a known barrier. By triangulating in this way, the analysis developed a credible representation of shared challenges.

Throughout coding and synthesis, the researcher was diligent about maintaining consistency and validity. The researcher also checked that every segment of text made sense with coding. The coding process resembled the approach described byMiles et al. (2014) – alternating between data and theory being generated. In situations where contradictions or divergent ideas arose, these were seen as potential sub-themes, or

"negative cases." As Shenton (2004) emphasises, for trustworthiness, triangulation of sources and methods strengthens credibility. In practice, an idea was only recognised as relevant if it was brought up in more than one source. For example, the interviewee concerns regarding *"no ownership after the sprint"* were only recognised as criteria in the tool when the same issue was present in both the survey and with the experts. Minor points mentioned by only one participant were treated cautiously and either merged with larger themes or noted for future exploration.

The final outcome of analysis was a draft of the "**Better Outcome" tool** consisting of several key criteria for evaluating sprint results. Each criterion (e.g. *Strategic Relevance*, *Feasibility*, *Stakeholder Buy-In*, *Public Value*, *Adaptatability*) was directly derived from the thematic findings. For example, the *Learning & Adaptation* criterion emerged from repeated references to the absence of formal reflection processes after sprints. The tool's language and format also underwent several iterations: language was simplified based on the input received from the experts, and criteria were added or collapsed based on feedback. For triangulation and confidence in these results, a form of **member checking** was applied. The researcher forwarded a summary of the findings and the evaluation criteria to one of the original interviewees and toone expert. Both members reviewed the summary and verified that the barriers and categories closely matched their ownexperience. They provided only a few minor corrections or suggestions, which were incorporated into the work after a discussion. This step aligned with Lincoln and Guba (1985) and Shenton (2004), providing reassurance that the interpretation accurately reflected the participants' realities.

In summary, the data analysis used neither a purely inductive or deductive approach, but rather some inductive openness and methodological rigor. The presence of systematic coding and cross checking allowed researcher to feel confident that the "Better Outcome" tool was empirically grounded in the collected evidence, and with the input of experts, was purposefully designed.

3.5 Ethical Considerations

As human subjects were involved in this research, the study was guided by ethical practice. Subjects were given comprehensive information on the research background, purpose and procedures involved, as well as their rights. In accordance with Creswell (2013), an

informed consent process was used. Both interviewees and expert participants were provided with a consent document that informed them their participation was voluntary and that they were free to withdraw from the study at any time. The consent document specified that personal identifiers would only be collected asto the participant's role and social background. Participants signed it, indicating they knew that their responses would be used in research and possible publication. As Creswell (2013) advises, the form “acknowledges that participants' rights will be protected during data collection” and that confidentiality measures are in place. At the start of each interview and expert session, the researcher reminded participants of these points.

In addition to the outstanding procedural restraints, the researcher was reflexive in practice to account for positionality and bias. As a participant observer aware of Policy Design Sprints, the researcher was mindful that his knowledge base of this field and assumptions raised during data interpretation may influence data iterpretation. To avoid it, neutrality was maintained in interactions: questions during interviews and walkthroughs were phrased openly, and the researcher refrained from sharing personal opinions or leading participants toward expected answers. The researcher enhanced the credibility and trustworthiness of the study by intentionally reflecting on what personal motivations could influence the research process (Lincoln & Guba, 1985; Patton, 2015).

The researcher also informed the interview participants and experts who participated in the walkthroughs of his dual role as both a facilitator of the Policy Design Sprints and a Researcher.The researcher provided background information to participants to address power imbalances and promote open discussion. By recognizing his own perspective and inviting feedback from colleagues, he aimed to focus on the participants' viewpoints in the study. This approach ensured that their insights were carefully considered throughout the research process.These reflexive practices functioned in conjunction with the formal ethical safeguards and further reinforced the credibility and trustworthiness of the findings (Creswell, 2014; Lincoln & Guba, 1985).

Confidentiality and Anonymity. All data were treated with respect to assuring privacy. Identifying aspects (i.e. names, organisations, projects) were removed from transcripts and notes. All participants when reporting results are referenced with generic codes (e.g. "Interviewee 1" or "Expert A") or by generic descriptions (e.g. " a city administrator"). Thisremoves any attribution of any specific quote to an individual or organisation.

Creswell (2014) describes confidentiality as an ethical imperative to guarantee values, and this was strictly followed. All raw data (audio files, transcripts, survey responses) are stored on a password-protected computer. Physical consent forms and notes are secured and available only to the researcher. The online survey was anonymous, and the researcher could not map responses to individual email addresses as they were not required.

Data Integrity. In order to facilitate authentic participation, the researcher took measures to ensure honest participation. Participants were advised that there are no "right" or "wrong" answers and that their honest feedback (even if critical of their organisation) was valued. No incentives were provided for participation to reduce the risk of deceitful answers. Shenton (2004) indicates that establishing rapport with participants and ensuring participant comfort with the research process are ethical strategies. Therefore, interviews consisted of friendly and respectful language on behalf of the researcher, and the researcher was careful to be mindful of power dynamics (e.g. not using leading questions). Interview participants were invited to skip any question they did not want to answer, and to end the interview at any time. Similarly, expert reviewers were aware that they could critique the tool without any restrictions; their role in the process was described as contributing to a professional peer review process.

Long-term Data Management. Related to ethical responsibility, there will not be long-term data retention. Practically speaking, electronic files will be stored securely for five years post publication, and then destroyed completely. At no time and in no place was additional data collected beyond what participants consented to. By adhering to these processes—consent, anonymisation, secured data storage - the research adhered to the values of respect for persons privacy and justice. The ethical safeguards ensured that participants could speak frankly about internal processes (for example, organizational shortcomings) without any risk of personal or professional consequences. In sum, ethical considerations were integrated into every stage of the research to protect participants and enhance the credibility of the findings.

4. RESULTS AND FINDINGS

This chapter presents the key results from the empirical study, including data collected via the survey, interviews, and expert walkthroughs (Section 4.1), followed by an integrated analysis of those results (Section 4.2). The analysis addresses the research questions concerning the post-sprint implementation gap, practitioners' notions of success, and the value of a structured evaluation tool. Finally, Section 4.3 offers a concise summary of the main findings, linking them back to theory and practice. All findings are expressed in a neutral academic tone from the single-author perspective.

4.1 Data Collection Results

Data were gathered through three complementary methods: an online survey of former policy sprint participants, in-depth interviews with sprint team leaders, and expert walkthroughs of the draft evaluation framework. The following subsections report the results from each of these data sources. Each is presented thematically with illustrative quotations where relevant.

4.1.1 Survey Results

Respondent profile and context: A total of 21 individuals completed the survey. All had participated in at least one government-led Policy Design Sprint in Latvia. The respondents have been working in a variety of settings (national agencies, municipal offices, and academia), ensuring diverse perspectives. Despite differences in context, clear patterns emerged regarding what happened to sprint outputs and how those outputs were (or were not) evaluated.

Implementation follow-up: The survey responses revealed a pronounced *implementation gap*. A majority of respondents indicated that the ideas or prototypes generated in their sprints had not been fully implemented in practice (83%). Instead, many reported that the sprint output was still "in progress" (49%) or "being discussed and refined," often with only partial execution (17%). In several cases, participants noted that although a concrete solution was produced, it received only lukewarm support from decision-makers and therefore stalled or was dismissed (17%). A few respondents openly admitted they are not involved in the next implementation steps, suggesting a lack of clear hand-off or ownership. As one survey participant commented, *".. whether it [sprint outcome] lives on*

depends on funding, political backing, and alignment with other regulations." This quote summarizes the reiterative theme that sprints generate enthusiasm and new ideas, however, often there are weak or no institutional processes to take those ideas further. In other words, many of the sprints result and conclude with ideas of promise that made it no further than an idea.

These findings echo concerns in the public innovation literature about the "last mile" problem (Ansell & Torfing, 2014b; van Buuren et al., 2020b). In both the scholarly literature and the survey data, the reasons cited for the post-sprint stagnation are similar: lack of political will or leadership support, shifts in priorities (such as changes of administration), resource constraints, or simple misalignment with existing legal or institutional frameworks. Such factors are frequently beyond the control of the sprint team and lie outside the short timeframe of the sprint itself. The survey responses consistently highlighted these barriers. For example, respondents noted that solutions fail simply because they contradict existing legal framework or require new legislation. In the absence of designated champions or funding, even strong ideas"die in a drawer" after the sprint. This confirms and extends prior research: Ansell and Torfing (2014b) and van Buuren et al. (2020a) argue that without explicit follow-up mechanisms, the impact of collaborative innovation methods will be blunted. The present data underscore that argument in the Latvian context. They suggest that concrete ideas often languish without integration into policy processes, highlighting the need for mechanisms to bridge the creative "laboratory" of the sprint and the conventional policymaking system.

Current evaluation practices: When survey respondents were asked whether their sprint's results had been evaluated, most said yes (86%) – but the nature of those evaluations was typically informal and unstructured (89% out of those who evaluated). In the absence of any standard framework, teams generally conducted only cursory debriefs or general discussions at the sprint's end, often without clear criteria or metrics. Only a minority of respondents answered that methods or tools were used to assess sprint outcomes (9,5%). In effect, if outcome evaluation occurred at all, it was ad hoc and left to each team's discretion. This finding confirms that structured evaluation is largely missing in current practice. It aligns with scholarly observations that formal post-project evaluation is often the "missing piece" in design-driven innovation (Ansell & Torfing, 2014b; Blomkamp, 2018): traditional policy evaluations usually occur much later and under strict

criteria that do not fit early-stage prototypes. Without any immediate assessment, teams lack systematic feedback or accountability for sprint outputs (Ansell & Torfing, 2014b).

Attitudes toward a structured evaluation tool: Despite the absence of prior frameworks, survey respondents expressed strong enthusiasm for the idea of a dedicated evaluation tool. Roughly 90% of survey respondents agreed that a structured evaluation process at the end of a sprint would have been useful. This is an important data point: far from rejecting the rationale for further process or perceiving evaluation as bureaucratic "red tape", practitioners overwhelmingly embraced the idea. Crucially, this positive attitude held regardless of whether a participant's last sprint had produced a successful outcome. Even those whose previous sprint outputs had been implemented were interested in having a structured tool for evaluating sprint outcomes. Only in two exceptional cases respondents answered that a new tool is not necessary.

Overall, the survey highlighted two key points: first, a gap in practice (sprint outcomes are rarely systematically evaluated, contributing to weak follow-through) and second, a clear demand for a solution (practitioners want a way to assess and learn from sprint outcomes). In short, the survey showed that teams recognized the problem of stalled ideas and see the need for constructive mechanisms to support accountability and implementation. All of this prepared the way for the qualitative data to provide a richer description and detail the indicatives' experience.

4.1.2 Interview Results

To explore these issues further, semi-structured interviews were conducted with four experienced policy sprint team leaders in Latvia. Each interviewee had led at least one sprint and could reflect on both successful and stalled cases. The interviews probed how they define sprint "success," what criteria they use to judge outcomes, what challenges they face in evaluation, and how they felt about a structured tool. The following themes emerged, illustrated with representative quotations (translated and anonymized).

Definitions of sprint success: All interviewees agreed that the ultimate measure of a sprint's success is its real-world impact. A sprint was considered "truly successful" only if its outcome led to a tangible improvement or was carried forward into practice. For example, one team lead (TeamLead B) succinctly stated, *"Primarily, success is if the result is carried forward and implemented – you manage to get support from the*

decision-makers." Another remarked that even the most innovative sprint is "largely a wasted effort" if the idea "dies in a drawer," echoing the sentiment that *"if nothing happens with the idea after, then it's hard to call it a victory."* This emphasis on implementation aligns with literature that calls for moving beyond ideation to actual value creation in public-sector innovation. In other words, learning and excitement matter, but they are only means to an end: the end must be change on the ground (Blomkamp, 2018; Ansell & Torfing, 2014b).

At the same time, the interviewees acknowledged that sprints often yield valuable intangible outcomes. Several noted that even if a specific solution isn't implemented, the process can still build team capacity, boost morale, generate new problem insights, foster cross-department collaboration, or spawn spin-off ideas. One respondent pointed out that sprints can create alignment and enthusiasm that pay off later, if not immediately. This more nuanced view matches scholarly work showing that design sprints can produce important collateral benefits (learning, networks, stakeholder buy-in, etc.) (Blomkamp, 2018; Ansell & Torfing, 2014b). However, all interviewees concurred that these intangibles, while useful, do not fully compensate for a lack of implementation. As one leader (TeamLead D) explained, *"It's great that people learn and get motivated, but in the public sector we ultimately need to see something change on the ground."* In short, practitioners see the process outcomes as important but subordinate to getting a solution adopted.

Criteria for evaluating outcomes: When asked how they assess or judge a sprint outcome's promise, the team leads described a remarkably consistent set of criteria. These criteria closely matched what emerged from the survey and alignwith innovation theory. Common factors included feasibility (can the solution actually be implemented given budget, regulatory, and political constraints?), stakeholder buy-in (is there a champion or broad support, especially among decision-makers and users?), public value (does it effectively solve a real problem and deliver societal benefit?), systemic fit (does it align with existing policies and institutions?), and sustainability (is the idea durable over time, or likely to be a one-off?). For instance, one team lead (TeamLead A) warned that *"we must ensure the solution doesn't break some other functioning system… sometimes a sprint team might not know all the connections."* This highlights the importance of regulatory and institutional fit. Others noted the necessity of having a project owner or champion; without

someone responsible, *"if a sprint result lacks a champion or fits poorly with priorities, it is unlikely to go anywhere."*

Importantly, these practitioner-defined criteria map closely onto multi-dimensional evaluation frameworks in the literature (Kimbell, 2016a; Tõnurist et al., 2017; De Vries et al., 2016). The interviewees implicitly referenced effectiveness, legitimacy, and adaptability, just as scholars do. For example, interviewees prioritised public value and problem alignment rather than novelty for its own sake. All four stressed feasibility above all – they immediately think about practical constraints, resources, and political support. They also placed high weight on stakeholder buy-in, mirroring calls for inclusive, co-produced solutions. In short, the criteria that practitioners described are multi-dimensional and in line with scholarly prescriptions for evaluating public-sector innovations (e.g., Kimbell, 2016a; Tõnurist et al., 2017). This congruence suggests that the theory has indeed captured what matters to the field.

Challenges in evaluating sprint outcomes: The team leads noted several obstacles to systematic evaluation. First, they confirmed that no established evaluation process exists in current practice. Teams simply have not been equipped with ready-made tools; evaluation, if done at all, has been informal. Second, even apart from the lack of a tool, the sprint format poses challenges. Outcomes are often at a low fidelity stage (sketches, rough prototypes), with uncertainty about next steps, making objective assessment hard. One lead remarked that the fast tempo and resource constraints of a sprint leave little time for reflection. Bias can creep in (e.g. teams and facilitators want their ideas to succeed,) and data to support feasibility (like cost estimates or user tests) may be sparse. In essence, the very structure of a design sprint — time-boxed and oriented toward creativity — has not naturally supported rigorous evaluation. Without any easy mechanism, teams have historically skipped formal assessment and moved directly toward whatever next step the organisation takes (or doesn't take).

Reactions to the proposed evaluation tool: Given these challenges, the interviewees were receptive to the idea of a structured evaluation framework (the "Better Outcome" tool) that they could use. Research participants broadly agreed, that a tool would address the identified gap as long as it acknowledged their constraints. In the interviews, there were many agreed points about what a tool should look like. Practitioners articulated that any evaluation process must be **lightweight and prompt** to not become an interruption to the

spirit and pace of the sprint. Ideally, the evaluation step could occur immediately after the sprint ends (for example, on the last day) so that it does not derail creativity. The tool should systematically guide teams through the key criteria (many of which they already consider informally), but in a streamlined way. Importantly, the interviewees strongly insisted that evaluation must not feel punitive or like a "grade" on the team'sperformance. Instead, it should be framed as a supportive aid – an opportunity to improve the idea and reflect on readiness, rather than a judgment.

Respondents also saw the potential of tool as a chance to fill an important gap in the sprint process. They expected it to help "build something to show to [their] boss," as one put it, by making implicit thinking explicit and providing justification for pursuing (or not pursuing) an idea. Another dimension of the tool surfaced by reframing it as a learning aid, it was seen as a way to illuminate reasons why the outcome stalled and serve as documented experience that can improve future chances.

In summary, all interviewees saw value in having a formal evaluation step. They all stressed, however, that in order to be adopted, the tool must be **flexible and supporting, and not burdensome**. Their responses were more descriptive than the survey results: they all painted a clear picture of the frustration with lack of follow through, but they were also optimistic about the worth and potential of design sprints and wanted to learn more. In effect, the interviews confirmed that practitioners recognize the implementation gap and lack of structure, yet they also articulated a clear sense of what matters in judging outcomes. Their insights provide a "reality check" on practical constraints (time pressure, data availability, cognitive bias) and on how a tool can be made viable. In doing so, the interviews bring a human and contextual richness to the results – illustrating the frustrations of seeing ideas stall, the hopeful creativity of sprints, and the practical wisdom about what must be in place for innovation to succeed.

4.1.3 Expert Walkthrough Results

The draft "Better Outcome" evaluation framework was then presented to three seasoned public-sector design sprint experts for validation and refinement. The experts are highly experienced facilitators of government innovation processes in Latvia. In guided sessions, the experts reviewed a prototype of the tool (a checklist of criteria with a simple rating

scale) and provided feedback on its clarity, relevance, usability, and anticipated effects. Their responses are summarised below.

Relevance and completeness of criteria: The experts agreed unanimously that the framework's criteria are highly relevant and cover the crucial dimensions of a sprint outcome. In their walk-through, each expert recognised in the tool's checklist the same considerations they themselves use informally. As one expert nodded and remarked, *"Yes, these are exactly the things I would want a team to think about before calling their idea finished."* In particular, the experts confirmed that the criteria on societal impact, feasibility, stakeholder/user validation, policy/system alignment, level of innovation, and sustainability were all essential. None identified any major aspect of sprint outcomes that was missing from the framework. In short, the tool's multi-dimensional criteria resonated strongly with the experts' own mental checklists. This endorsement provides face validity: it indicates that the evaluation dimensions derived from the literature and fieldwork capture what truly matters in practice. One expert observed that in the sprints she had facilitated, teams often considered these questions informally (for example, asking "Will the boss approve this?" or "Did we check the legal aspects?"), but doing so in a structured manner is rare. The experts concurred that a formal checklist would help ensure"nothing falls through the cracks," given the time pressure in sprints. The close convergence between the tool's content and the experts' expectations confirms that the framework is conceptually well-founded.

Usability and process integration: The experts also reflected on how the tool would fit within the sprint process. They emphasised that it should be lightweight and easy to use. During the walkthroughs, they generally found the tool's format (brief criteria plus rating guides) intuitive and feasible to complete within a sprint context. Several experts highlighted the idea of integrating evaluation with the sprint timeline: for instance, introducing the criteria at the sprint kickoff could orient teams toward desired outcomes, while using the tool as a self-evaluation at the end could solidify learning. One insightful comment (echoing what the team leads had suggested) was that defining evaluation criteria at the start of the sprint could actually shape the process: *"If the team knows on day one what a 'good outcome' should look like, they might design their solutions more intentionally to meet those criteria."* Thus, evaluation would not be an afterthought but a guiding thread through the sprint.

Refinements and final validations: From the experts' feedback, the framework was adjusted slightly. For example, the experts said to include the prompt regarding "ownership," so that someone is responsible for follow-up, and to specify or clarify a couple of the criterion definitions to be clear in their meaning. The experts endorsed the tool as ready for a trial with those very minor adjustments. They viewed it as a validated instrument: all three stated in effect that *"we would use this"* in our sprints. Their enthusiastic endorsements lend credibility to the tool's viability. In sum, the expert review confirmed that the framework aligns with practitioner needs and is practical in a real-world context.

Conclusion of data collection: In conclusion, the combined data collection phase found consistent evidence of an implementation gap in policy sprints, a shared set of success criteria among practitioners, and strong practitioner support for a structured evaluation framework. The survey and interview data established the "what" (sprints stall without follow-up, people care about certain evaluation dimensions, there is demand for a tool), and the expert walkthrough confirmed the "how" (the tool's content and format are appropriate). These results have been used to finalise the "Better Outcome" framework, preparing it for deployment in actual sprints (see Chapter 5).

4.2 Data Analysis

This section synthesizes the findings from all data sources and interprets them in light of the research questions and theory. Three major themes structure the analysis: (1) the implementation gap and need for follow-up, (2) the alignment of practitioner-defined success criteria with theoretical frameworks, and (3) practitioners' support for a structured evaluation tool and its expected impact.

4.2.1 The Implementation Gap and Need for Follow-up

A clear, fundamental finding is that government Policy Design Sprints frequently suffer from a post-sprint *implementation gap*. Both the survey data and interview narratives consistently show that teams often generate viable concepts, yet those concepts "stall" instead of becoming implemented solutions. In effect, sprints produce innovation *in vitro* but have trouble translating it *in vivo*. This echoes documented concerns in the literature: government innovation labs and design sprints are known to struggle with the "last mile" of policy change (Olejniczak et al., 2020a; van Buuren et al., 2020a). Data provided

concrete evidence of this phenomenon in the Latvian context. Numerous respondents reported their sprint outputs lingering without adoption or only partially realised months later.

The causes of this gap became evident in the responses. Practitioners and survey respondents identified factors that are typically outside the sprint team's control: lack of political will or changes in political leadership; insufficient budget or staffing to take the idea forward; misalignment with existing laws or institutional norms; and absence of a designated champion to drive the project. These reasons align with prior research on public innovation implementation, highlighting political support and institutional fit as critical success factors. For example, Ansell and Torfing (2014a) emphasise that collaborative innovation requires "metagovernance" – deliberate attention to not just idea generation but also evaluation and integration of new ideas. Without someone formally tasked with follow-up, good ideas often "fall through the cracks" after the sprint.

In the words of one respondent, a sprint team may achieve "something to show," but if the organisation's higher-ups do not commit to the idea, the work stops there. This situation is a classic case of bureaucratic momentum dissipating once the sprint ends. The structured analysis suggests that the problem is not a lack of creativity or community engagement (which sprints mostly provide), but a lack of mechanisms to carry those innovations into standard practice. In other words, the bridge between the creative sprint and the mainstream policy process is weak. The participants themselves implicitly echoed this assessment: many noted that in practice, *"no one is formally tasked with carrying sprint ideas forward."*

The implication is that deliberate follow-up is needed. All the data converge on the idea that sprints should not be one-off events whose outputs simply float away; instead, there must be integrated steps that connect the sprint back into the policy system. Introducing a structured evaluation and follow-up step (as this research does) could serve as one such mechanism. By institutionalising a more immediate assessment of the idea's readiness and barriers, the sprint methodology could incorporate accountability and continuity. This would begin to bridge the known "integration gap" (Christiansen, 2014) - the gap between the experimental creativity of sprints and the practical need for governance in implementation. The evidence suggests there is a real and substantive gap in implementation processes of design sprints, pointing to the need to develop mechanisms

(for example, evaluation frameworks and dedicated champions) to avoid losing viable ideas.

4.2.2 Alignment of Practitioner Criteria with Theoretical Frameworks

A second major insight is the strong correspondence between the success criteria that practitioners use and the dimensions emphasized in academic evaluation frameworks. Both the survey and the interviews revealed that practitioners think about sprint outcomes in rich, multi-dimensional terms. They consider feasibility (budget, legal, and political feasibility), stakeholder support (buy-in from decision-makers and users), public value (real impact on the target problem), systemic fit, innovativeness, sustainability, and clarity of the solution. These dimensions, identified independently by the data collection, closely mirror what innovation scholars have advocated.

For instance, the literature on public-sector innovation stresses that evaluation should go *beyond* narrow cost-benefit and include broader concepts like public value and legitimacy (Kimbell, 2016b). It also emphasises adaptability and systemic alignment as key factors (Tõnurist et al., 2017). Remarkably, every one of these concerns emerged organically from the practitioners' descriptions. This alignment validates both theory and practice. From the theoretical side, it suggests that academic frameworks have successfully captured what truly matters "on the ground." From the practitioners' side, it suggests that when a formal tool is introduced based on these frameworks, it will not impose unfamiliar criteria; instead, it will simply codify what teams already care about.

Indeed, the Better Outcome tool's criteria were derived from a combination of literature and early field input, and the expert validation showed that experienced facilitators found them comprehensive and transparent. The fact that both the practitioners (through their own words) and the experts recognised the exact evaluation dimensions strengthens confidence in the evaluation approach. As one expert noted, teams often informally ask questions like “Will the boss approve this?” or “Have we checked the budget?” – the tool simply makes those questions explicit.

In sum, this study finds a close correspondence between practitioner-defined success measures and those found in the literature (De Vries et al., 2016). Whether called “public value,” "legitimacy,” or "feasibility,” the key factors are shared. This convergence is encouraging: it implies that introducing a structured evaluation framework will reinforce

existing priorities rather than add alien concepts. Practitioners intuitively value these broad criteria already, and the framework merely provides a systematic way to apply them. In practical terms, it means that the evaluation framework has strong content validity: the criteria on which sprint outcomes will be judged are ones that practitioners themselves endorsed, giving the tool relevance and credibility.

4.2.3 Support for and Impact of a Structured Evaluation Tool

The third theme is practitioners' receptiveness to a structured evaluation tool and the expected benefits. The data leave no doubt on this point: there is broad support among public-sector innovators for such a tool, and they believe it will make a positive difference. This addresses the research question about stakeholder acceptance head-on. Approximately 90% of survey respondents viewed a dedicated outcome evaluation framework as beneficial. Moreover, every interviewee and expert participant uniformly supported the idea, seeing it not as bureaucratic overhead but as something helpful and evenempowering.

This enthusiasm is noteworthy. One might have predicted that creative teams would resist formal evaluation as stifling, yet the opposite occurred. According to the analysis, there are three main factors that explain this support. First, there is clear frustration with the status quo: participants are well aware that sprints often end in unimplemented ideas and wasted potential, so they are hungry for any solution that improves follow-through. They explicitly said they want to learn from each project and document what happened. Second, the content of the tool resonates deeply with their own values and concerns. Because the framework's criteria match what they already think is important, it does not feel alien. It feels like formalizing the checklist they have in their heads. Third, there is a desire for success and recognition: practitioners want to showcase their winning ideas and understand why others fail. If a tool can help them articulate and demonstrate the value of their work, they welcome it.

The experts implied that the tool could provide additional and real value. The proposed evaluation framework, used in and with the final sprint report, would give the team a clearer picture of the strengths and weaknesses of their solutions. For example, in including a section for stakeholder validation or legal fit to the process, teams may be able to build this requirement into their design habits. Experts anticipated that this would lead to more stakeholder-conscious design decisions and smoother hand-offs to implementers. In

theory-of-change (Weiss, 1995) terms, introducing the evaluation step should lead to more thoroughly vetted sprint concepts (with known risks and mitigation plans), which in turn should increase the likelihood of those concepts being successfully adopted. In other words, the tool is expected to improve the "signal-to-noise" ratio of sprint outcomes by addressing common failure points.

Although this study did not track long-term outcomes, the qualitative evidence strongly suggests that projects going through the "Better Outcome" process would command more confidence from teams and decision-makers. Participants believed that knowing an idea had passed a rigorous, transparent check would make it easier to justify follow-up resources. This reinforces the real-world significance of the tool: it is not an intellectual exercise, but something that participants believe will influence a real-world outcome.

There are some caveats to consider. Both practitioners and experts emphasised that institutions should be in the right organisational context for the tool to be of value. Organisational support and a learning culture will be essential. If the evaluation is introduced as a punitive "grading" mechanism or if there is a culture of blame, it could backfire. However, the trial feedback revealed no inherent resistance; rather, it highlighted that proper introduction and supportive framing are key. Assuming a positive stance, the expected benefits (better outcomes, knowledge transfer, team learning) appear to outweigh the modest additional effort of using the tool. In essence, participants see the evaluation framework as directly addressing existing pain points in the sprint process and are inclined to adopt it.

Cross-cutting analysis: Bringing these themes together, three overarching insights emerge. (a) Policy Design Sprints are effective at generating creative concepts but often face a serious implementation gap; deliberate follow-up mechanisms are needed to bridge the sprint outputs to policy impact. (b) Practitioners' criteria for success are broad and multi-dimensional, closely mirroring scholarly evaluation frameworks; this validates both the relevance of theoretical criteria and the decision to base the tool on them. (c) There is strong bottom-up support for using a structured evaluation framework, with practitioners expecting it to improve outcomes rather than hinder innovation. Together, these findings paint a picture of a field that is ready to evolve: teams recognize the need for more systematic evaluation and learning, and the proposed "Better Outcome" framework offers a

timely response. The research has effectively closed the loop: insights from the survey and interviews were translated into the tool, which expert feedback then confirmed and refined.

4.3. Summary of Findings

The most significant findings of the study are summarised below, and the implications of the results and analysis are summarised. These findings directly address the study's research questions and demonstrate how to think about the contributions to theory and practice.

1. **Implementation gap in policy sprints:** While public-sector design sprints can produce potentially valuable ideas, there is a significant implementation gap and lack of structured approach to encourage post-sprint follow-up. Sprint teams often realise that their solutions stall when the design sprint is over. This is frequently due to unclear ownership, missing political support or resources, or simply not fitting with the institution or its strategy. In short, many sprint outputs do not translate into real-world change. This underscores the need for deliberate post-sprint integration mechanisms (Ansell & Torfing, 2014b; van Buuren et al., 2020a). Without such mechanisms, creative ideation alone cannot guarantee impact.
2. **Multi-dimensional success criteria:** Practitioners define sprint “success” in multi-dimensional terms that strongly align with theoretical evaluation frameworks. Rather than judging success solely on immediate deliverables, practitioners emphasized criteria such as feasibility (budget, legal, political practicability), stakeholder buy-in (support from decision-makers and users), public value (meaningful problem-solving impact), sustainability, and systemic fit. These priorities mirror what scholars advocate for evaluating public-sector innovations (e.g., Kimbell, 2016b; Tõnurist et al., 2017; De Vries et al., 2016). The alignment means that the evaluation criteria derived from the literature were validated by real-world perspectives. In practice, the team’s checklist and the academic checklist turn out to be one and the same, lending confidence to the chosen evaluation approach.
3. **Appetite for structured evaluation:** There is a strong appetite among public-sector innovators for a structured tool to evaluate sprint outcomes, and it is

seen as addressing an important gap. Rather than resisting additional evaluation steps, practitioners generally welcomed the idea. Roughly 90% of surveyed participants favored using a dedicated outcome evaluation framework, and all interviewees and experts uniformly supported the concept. They do not view such a tool as bureaucratic red tape, but as a helpful aid to ensure impact. This bottom-up support reflects a broader shift toward evidence-based innovation practice (De Vries et al., 2016): practitioners want to learn from each project and improve their chances of success. In effect, teams are ready and willing to incorporate more rigorous evaluation into their process.

4. **Validation of the Better Outcome framework:** The "Better Outcome" evaluation framework, which was piloted through expert walkthroughs, is demonstrated as a viable way to improve the evaluation of sprint outcomes. Experts found the tool's evaluation criteria to be comprehensive and the process manageable in a way that is necessary in a sprint context. They endorsed its usefulness and suggested minor refinements (for example, adding a prompt to assign ownership of follow-up and clarifying criterion definitions) to enhance it. With those refinements, the framework is ready for implementation: it is easy to use, aligns with practitioner values, and is expected to yield better project outcomes through improved clarity, knowledge transfer, and follow-through. This validates the research approach of translating the findings into a practical artefact. The framework develops operational elements of the study insights and represents a tangible contribution to better practices in design sprints.

In summary, research results highlight a clear need and opportunity for the integration of systematic evaluation of outcomes into Policy Design Sprints. Doing so can bridge the gap between creative ideation and practical implementation, ensuring that the energy and ideas generated in sprints lead to real public value. The convergence of practitioner enthusiasm and theoretical support for this approach strengthens the case that introducing structured evaluation (such as via the Better Outcome tool) can enhance both the process and outcomes of public-sector innovation initiatives.

5. DESIGN WORKS

The development of the "Better Outcome" tool was a design process that occurred in multiple iterations, influenced by the opportunities provided by rich insights gathered from surveys, semi-structured interviews, and experts' walkthroughs of the tools prototype. The whole purpose of the prototype was to address the gap identified in the evaluation of the Policy Design Sprint by providing facilitators and policy sprint teams a structured, efficient and reflective way to evaluate outcomes of the sprints they are involved in. This section will describe the iterative development and finished form of the prototype, elaborating on the conceptual rationale, interaction design and evaluative aspect of the tool.

The tool is designed as an instrument that is generally used by facilitator in a particular sequence to gain the most value of the right timing. Its process is graphically visualised in the “Suggested use of “Better Outcome” tool during the Policy Design Sprint” (Figure 1.)

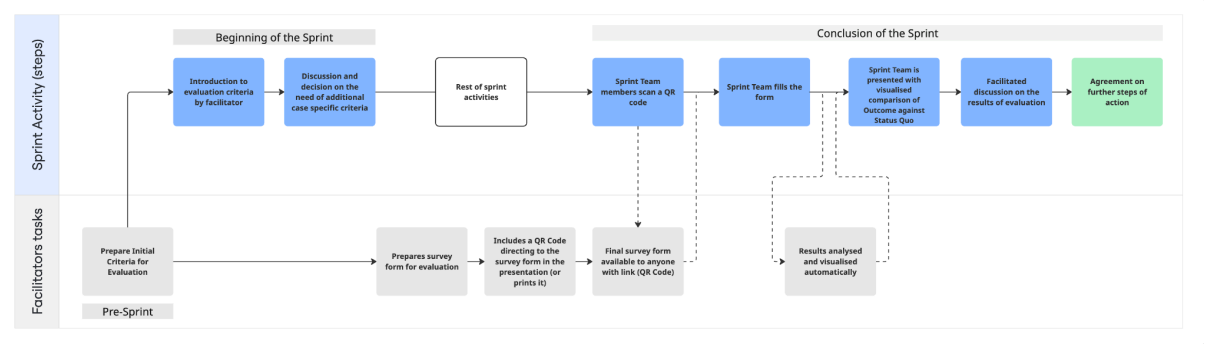

Figure 1. “Suggested use of “Better Outcome” tool during the Policy Design Sprint”

5.1. Initial Ideation and Iterative Development

The conception of the “Better Outcome” tool was a consequence of extensive research about existing evaluation frameworks and methodologies, recognising that there were no streamlined evaluation systems to evaluate the outcomes of Policy Design Sprint (Bason, 2017; Stickdorn et al., 2018). Initial ideation focused on simplicity, ease of use, and rapid feedback central to the time sensitive nature of policy sprints. Primary ideas identified in the early development often revolved around aspects that had to be accessible via mobile technology, and have minimal administration burden, and quantitative feedback mechanisms so facilitators and participants could interpret individual and collective outcomes instantly.

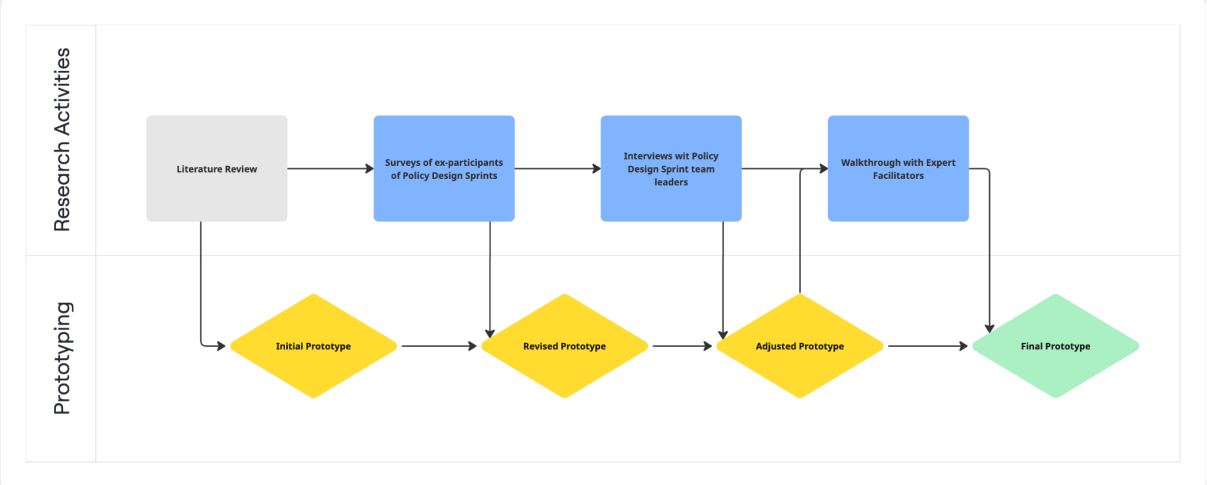

Figure 2. "Diagram "Iterative development of prototype""

Initial concepts were refined using insights from the data obtained from the survey and semi-structured interviews with sprint team leaders (see "Visualised representation of the sprint team's self assesment on feasibiilty" (Figure 2)). Another layer was implemented after testing the tool prototype with expert facilitators. These conversations highlighted the importance of clear criteria concerning the feasibility, alignment with systemic conditions, and practical impact of sprint outcomes (Creswell, 2013).

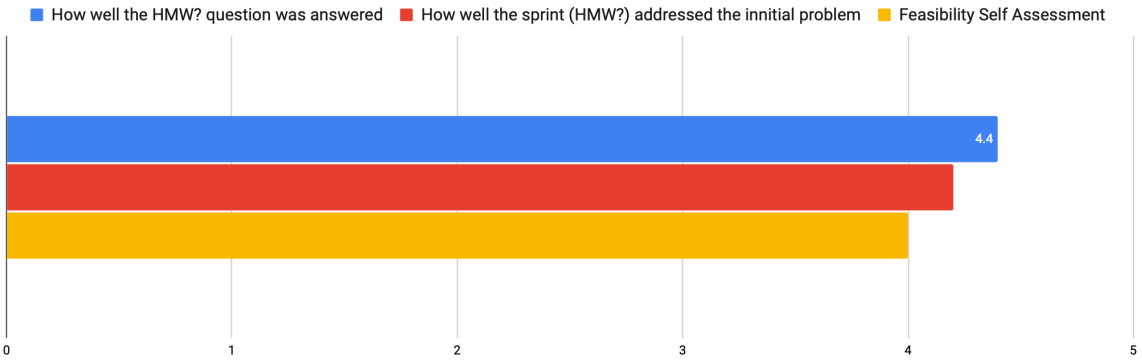

Figure 3. "Visualised representation of the sprint team's self assesment on feasibiilty"

The early prototype contained initial sets of criteria based on academic literature and field observations, subsequently refined by insights from the interviews with Policy Design Sprint team leaders (resulting in the addition of a self-assessment section of feasibility and compatibility with given task - see "Visualised representation of the sprint team's self assesment on feasibiilty" (Figure 3.)) and finally through expert walkthroughs another field was added to the self-assesment section (who is going to lead the follow-up?).

5.2 QR Code and Digital Accessibility

In order to facilitate easy access and provide the easiest way for the evaluation tool to be integrated into the sprint process, the prototype was developed from the outset as being accessible online through a hyperlink and a corresponding Quick Response (QR) code to grant easy access for participants. The design decision was made to reduce the barriers to participants' engagement, aligning with a user-centred design literature (Norman, 2013). Facilitators display the QR code in a physical or digital presentation to provide instantaneous access via participants' devices, ensuring the evaluation process can maintain fluidity without disrupting the momentum of the sprint dynamics or the collaborative energy.

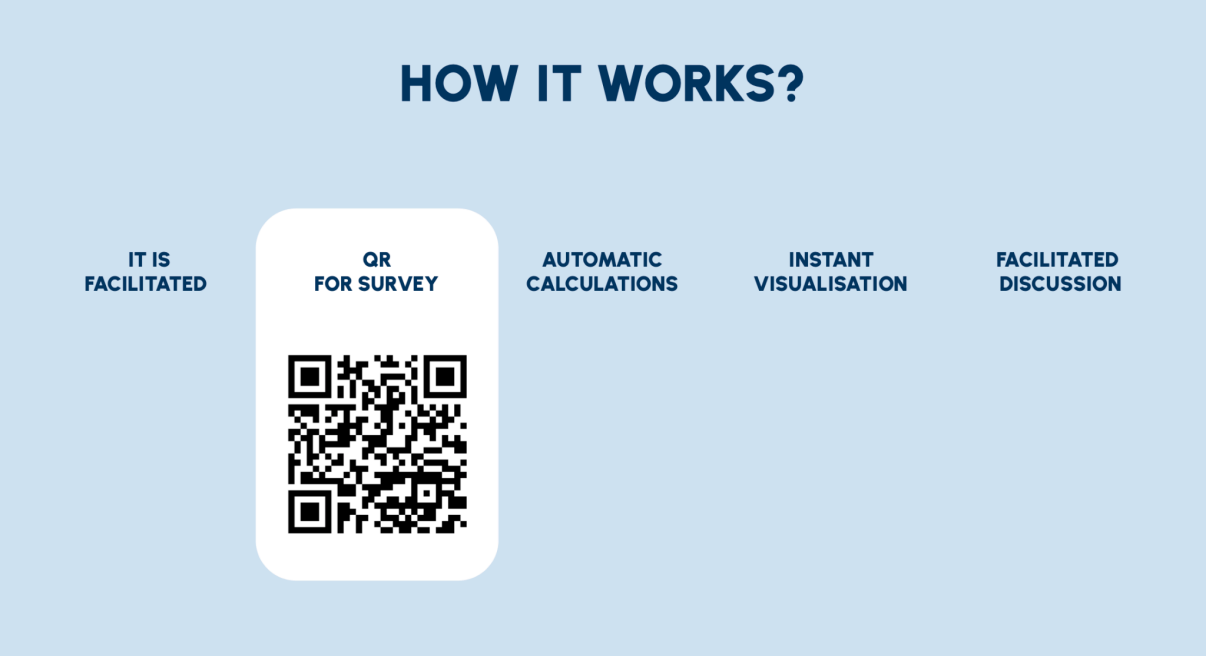

Figure 4. "Ilustrative slide with QR code from presentation for Expert Walkthroughs"

5.3 Form Structure and Likert Scale Integration

The Better Outcome form made available through the QR code was designed entirely around a Likert Scale. The reason for using the Likert Scale allows for a high level of explanation around eliciting nuances of levels of agreement or disagreement quickly and efficiently. Likert scales are particularly effective at obtaining clear and succinct participant feedback, without the complications and potential confusion of qualitative open-ended questions, thus improving usability and speed of data collection (Fowler, 2014).

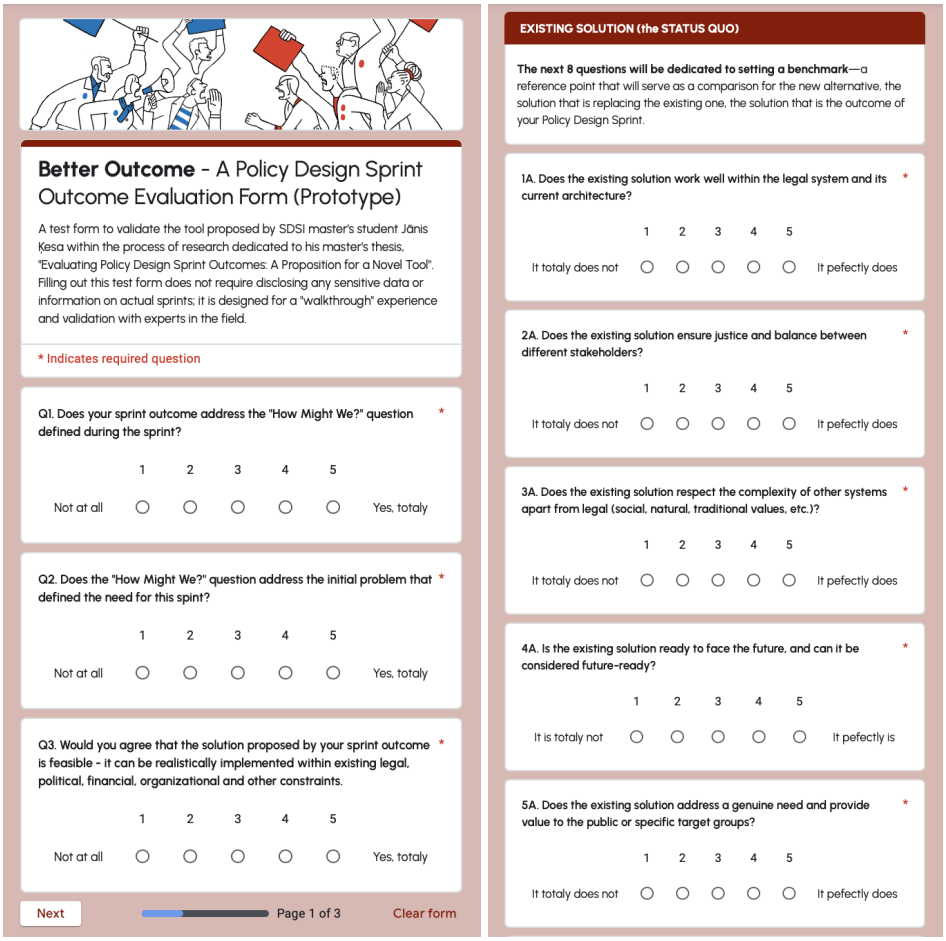

Figure 5. "Ilustrative example of Likert Scale integration in Survey phase of "Better Outcome" tool"

The tool has a clearly structured set of questions divided into three groups. The first set of questions are non-comparative self-assesment questions dealing with feasibility and role division for the follow up. The following two groups are dedicated to comparative evaluation, Existing Solution (Status Quo) and Proposed Solution (Sprint Outcome) questions, respectively. Each group contains eight questions that can be addressed using the exact same evaluation dimensions derived from existing Service Design, Systems Thinking, and Policy evaluation frameworks (Stickdorn et al., 2018; Thaler & Sunstein, 2008).

5.4 Comparative Evaluation and Benchmarking Approach

As stated previously, the "Better Outcome" tool is comparative and benchmarking-based. It addresses the critical need to measure and demonstrate the value of the proposed solution against a defined baseline or status quo (McGann et al., 2018; Mercure et al., 2021b).

Participants will first measure the current policy or service against legal compatibility, justice among stakeholders, systemic coherence, future readiness, public value, understandability, behavioural insights, and administrative efficiency. Following this assessment, participants will use the same measuring framework and answer the same questions to review the solution that emerged from the sprint. In this regard, the direct comparison allows the measure to be transparent, organised, and fair, while providing a precise identification of each of the proposal's strengths and possible weaknesses.

5.5 Evaluation Dimensions and Theoretical Justification

Each of the selected evaluation criteria included in the “Better Outcome” tool reflects solid theoretical foundation drawing from Service Design, Systems Thinking, behavoural Economics, Legal Design, and Futures Thinking. Each question aligns explicitly with evaluation dimensions proven critical in determining the practicality, acceptability, and impact of policy innovations (Thaler & Sunstein, 2008; Weber & Rohracher, 2012; Edquist, 2020).

For instance, the evaluation criteria for feasibility and systems coherence stem from Systems Thinking, emphasising that effective solutions must respect and integrate into existing institutional and societal frameworks (Jones & Bowes, 2017). Evaluation of the Behavioural Insights dimension assesses whether the proposed solution effectively considers user behaviour, aligning with insights from Behavioural Economics, which highlight the importance of user engagement and adoption (Thaler & Sunstein, 2008). Likewise, the evaluation for future-ready assesses elements from Futures Thinking methodologies focusing on the proposed solutions' resilience and ability to adapt to evolving conditions (Conway et al., 2018).

5.6 Radar graph visualisation

A critical component of the “Better Outcome” tool is the visual representation of evaluation outcomes through a radar graph, automatically generated based on participant responses. The “Better Outcome” tool employs an automatically generated radar graph to visually represent evaluation results derived from participant feedback. This graph serves as a critical component of the tool's functionality. The radar graph is an excellent format because it can quickly help facilitate the visual comparison across multiple evaluation criteria through a single visual; offering transparency to the strengths and weaknesses

comparing the proposed solution versus status quo. This visualization allows facilitators and decision makers to easily make sense of complex evaluative data, so they can have informed conversations and make decisions regarding the adoption or further development of sprint-generated outcomes.

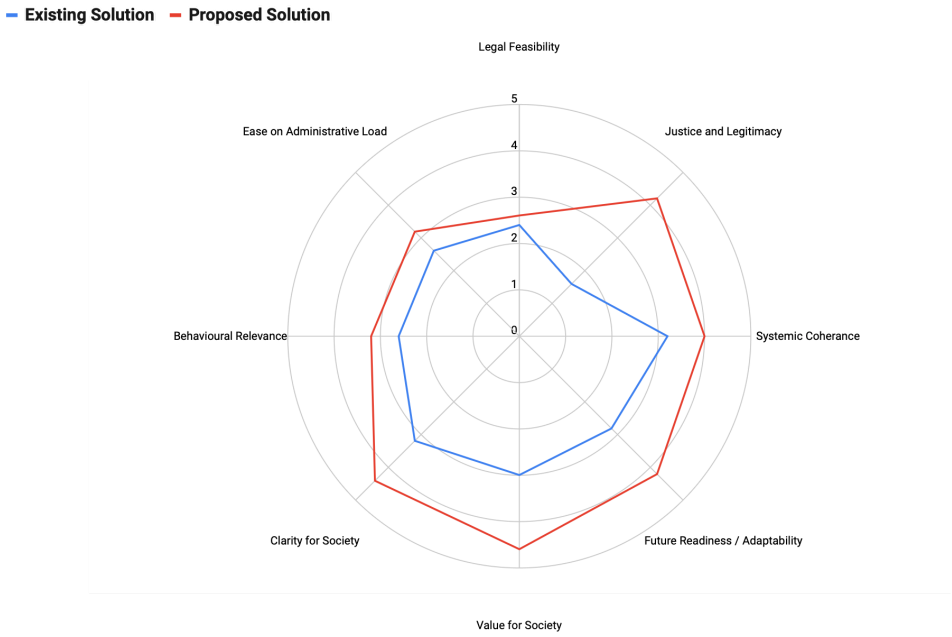

Figure 6. "Visualised representation of the outcome against the benchmark using radar graph"

5.7 Usability and Practical Validation

The usability of the prototype was extensively assessed through expert walkthrough sessions, where three, highly experienced policy sprint facilitators engaged with the tool, applying it to hypothetical sprint scenarios. The feedback from the experts valued the high usability, largely as a result of employing the very easy-to-use likert-scale system, and the intuitive comparative layout. Experts also valued the radar graph as immediate visual feedback would prompt meaningful conversations within sprint teams and decision makers (Nielsen, 1993).

Based on expert feedback, minor adjustments to the prototype of "Better Outcome" were incorporated, most of which related to the language and phrasing of specific evaluation questions for clarity and to limit ambiguity.

5.8 Ethical Considerations in Design

Ethical considerations were successfully addressed through the design and testing phases and reinforced key aspects of anonymity, voluntary participation, and confidentiality throughout the design and testing phases. The prototype of “Better Outcome” was designed considering ethical aspects and to avoid sensitive information, focusing entirely on sprint outcome evaluation criteria rather than organizational specifics or personal detail. This respectful approach aligns with ethical standards of research and professional practice, providing the participants with a sense of security while providing honest and constructive feedback (Lincoln & Guba, 1985).

5.9 Design Contributions

The “Better Outcome” tool is a valuable step forward in how Policy Design Sprint outcomes are evaluated. It takes complex set of criteria and presents the evaluation data in a straightforward, easy-to-understand format that's visually appealing. This approach helps bridging a significant gap in how policy innovation through design sprint methodology is currently done. With its clear comparative framework and ability to provide instant visual feedback, the tool stands out as both an evaluation method and resource for communicating the outcomes, learning the reasons of success and failures while contributing to the “institutional memory”. It's designed to improve understanding, involve stakeholders more effectively, and foster ongoing improvement in public sector innovation by use of facilitated co-creation in a form of Policy Design Sprint.

Overall, the research and design process including the resulting tool underscore the importance of interdisciplinary integration, practical usability, and responsively iterative design, contributing meaningfully to both academic discourse and practice of evaluation in context with Policy Design Sprints.

6. DISCUSSION AND IMPLEMENTATIONS

This chapter seeks to interpret the central results of the research, linking them to the theoretical underpinnings articulated in the literature review. It addresses the meaning behind the implementation gap, the relationship between practitioner knowledge and academic theory, the importance of including a variety of analytical viewpoints in the evaluation process, and the motivations underlying practitioners' willingness to engage with formal assessment. The discussion will clarify the theoretical and practical contributions of this research. Detailed instructions will be provided for using the "Better Outcome" evaluation tool, including practical implications and challenges identified in the study. The chapter alsoidentifies limits based on this study's scope and methodology and concludes with some promising directions for future research.

6.1. Discussion

The empirical results present important insights into the realities of using Policy Design Sprints in government contexts. Framing the empirical results in relation to the literature enables greater insights into challenges and opportunities related to evaluating and implementing innovation generated through Policy Design Sprint. This section explores four key areas: the persistent implementation gap, the challenge of defining "better" outcomes, the need for integrated and diverse evaluation perspectives, as well as the rationale behind the practitioner support for structured assesment.

6.1.1. The Persistence of the Implementation Gap

A central finding from the empirical research was confirming a major gap between generating ideas through Policy Design Sprints to implementing those ideas into specific policy activity or service improvements in the Latvian public sector (Analysis 4.2.1). Survey data indicated that a remarkable 83% of overall sprint outputs were not implemented, with almost half (43%) vaguely noted as still "in progress," and almost a fifth (17%) said to be stalling entirely or being rejected (Finding 4.1.1). Interview data offered reinforced this finding, with team leaders indicating that good ideas often "die in a drawer" (Finding 4.1.2), rendering the sprint effort potentially "wasted" if no tangible change follows. This evidence strongly suggests that while sprints may generate collaboration and new ideas, they often struggle with the critical "last mile" of turning the

sprint outcomes into practical applications.

This observed phenomenon speaks to many established concepts in the literature within policy studies and public administration. It reflects the well-established "implementation deficit," in which public policies that have been fullydesigned on paper run into real challenges during execution (Howlett, 2018; Peters, 2018b). Moreover, the difficulties mentioned resonate with well-known barriers to diffusing innovations in bureaucratic public sector environments that areoften prioritising stability and predictability over experimentation (Mulgan & Albury, 2003; Bason, 2018; Kimbell & Bailey, 2017).

The factors contributing to this implementation gap, according to survey respondents and interviewees, point towards challenges above and beyond simple project management issues. Barriers including lack of sustained political will, unexpected leadership or priority change, inadequate funding or staff to support implementation, not fitting into existing legal and regulatory frameworks, and particularly a lack of project champions or owners were all articulated (Findings 4.1.1, 4.1.2, Analysis 4.2.1). Barriers of this nature are not just logistical inconveniences, they are institutional challenges that are intrinsic to establishing new, agile approaches to traditional governance processes.

When taken together then, these elements suggest that the implementation gap may be a symptom of a deeper systemic disconnect. Policy Design Sprints, a methodology borrowed from private sector contexts focusing on product development (Knapp et al., 2016), emphasise speed, user-focus, and iterative prototyping to generate innovative, lean plans. While valuable, this approach may face lack of essential means to navigate the complex, often path-dependent, and politically sensitive landscape of public policymaking. The sprint may function as a temporary bounded space for creativity, however, the outputs of the sprint might face considerable friction when they engage the established work routines, power relations, resource allocation, and legal constrains of the broader administrative system. The fact that implementation flatters due to factors almost exclusively out of the sprint team's control is indicative of this disconnect. This implies that the sprint framework, as it is being conducted publicly in many instances, sits in the governmental system but it has not been sufficiently *integrated into* its core decision-making and operational pathways. Therefore addressing the implementation gap requires more than just refining sprint methodology; it necessitates a more fundamental consideration of how governmental institutions absorb innovations emerging from agile co-creative design

processes in terms of authorisation, resource allocation, administration, and ultimately impementation. The "Better Outcome" tool, as proposed, aims to be a bridging mechanism by imposing an early review and assessment of these systemic alignment and feasibility factors, however, the viability of the tool will ultimately depend on the receptiveness and adaptability of the surrounding organisational environment (Ansell & Torfing, 2014b; van Buuren et al., 2020a).

6.1.2. Defining and Evaluating "Better" Outcomes

Another significant finding is the remarkable intersection between the criteria practitioners implicitly use without a provided framework to assess sprint outcomes and the multi-dimensional frameworks proposed in academic literature (Analysis 4.2.2). Despite the prevalent lack of formal evaluation tools (Finding 4.1.1), interviewees consistently articulated a sophisticated understanding of what constitutes a promising or successful sprint result. Their assessments considered factors such as practical feasibility (including budgetary, legal, technical, and political constraints), the degree of stakeholder buy-in (especially from crucial decision-makers and end-users), the potential for generating public value (solving the targeted problem effectively and delivering societal benefits), alignment with existing policies and institutional structures (systemic fit), and the potential for long-term sustainability (Finding 4.1.2).

These practitioner-defined dimensions map closely onto the criteria highlighted in the literature review as essential for evaluating public sector innovations. Concepts such as Public Value (Moore, 1995), Democratic Legitimacy and Stakeholder Acceptance (Clarke & Craft, 2019; Lewis et al., 2020), Feasibility and Strategic Alignment (Howlett, 2018; Peters, 2018b), and User-Centric Quality and Experience (Norman, 2013; Blomkamp, 2018) were all implicitly or explicitly reflected in the practitioners' discourse. The interview analysis noted that team leads intuitively weighed effectiveness, legitimacy, and adaptability when discussing past sprints (Analysis 4.2.2). This strong alignment serves a dual validation function. On one hand, it confirms that the theoretical constructs developed by scholars accurately capture the complexities and considerations relevant to practitioners working "on the ground." On the other hand, it demonstrates that public sector innovators possess a rich, contextually grounded understanding of the multifaceted nature of policy success, even if this understanding is not always formally articulated or measured.

This convergence suggests that the "Better Outcome" tool, which was designed drawing upon both the literature and the empirical findings (Chapter 4 description), is built upon a foundation of criteria that are already recognised as relevant and important by its intended users. The positive reception of the tool's criteria during expert walkthroughs further supports this, with facilitators confirming that the framework covered the essential dimensions they would expect teams to consider (Finding 4.1.3).

This leads to a further reflection: the tacit knowledge held by experienced public servants and sprint participants represents a valuable, yet currently underutilised, resource for evaluation. Practitioners are already making judgments based on feasibility, political winds, user needs, and systemic constraints, but because these assessments often remain informal and undocumented (Finding 4.1.1), the insights are not systematically captured, shared across teams, or leveraged for organisational learning. The "Better Outcome" tool, therefore, does more than just impose an external assessment structure; it provides a mechanism for surfacing, articulating, and formalising this existing practical wisdom.By prompting teams to explicitly discuss and rate dimensions like stakeholder buy-in or systemic fit, the tool encourages the conversion of tacit understanding into explicit knowledge that can inform immediate decisions and contribute to a collective evidence base over time. Its value extends beyond just assessment. It encompasses knowledge management and organisational learning both important aspects contributing to the skills of public sector teams to improve the quality and viability of design-led innovations.

6.1.3. The Need for Integrated Perspectives

The criteria that practitioners highlighted, especially those relating to systemic fit, sustainability and stakeholder acceptance (Finding 4.1.2), implicitly suggest that it is important to assess the outcomes of a sprint using perspectives that are not limited to the features of the prototype alone. The cautioning comment from one participant about ensuring that a new solution does not, in effect, bring about the unwelcome outcome of “break some other functioning system” (Finding 4.1.2) illustrates that the participant was mindful of broader interdependencies. This also resonates with the development of the literature review section that provides a rationale for employing Systems Thinking, Behavioural Insights and Futures Thinking into an evaluation of policy innovations (Lit Review 2.4).

Traditional evaluation tends to concentrate primarily on the functionality of the prototype (typical outcome of sprint) or the immediate feedback from users involved in testing it. However, the demands of the public-sector complexities typically call for view through a wider lense. Systems Thinking argues for an examination of how a new solution fits withthe interconnected megasystem of institutions, regulations, power dynamics, and stakeholder networks (Klein Woolthuis et al., 2005; Pierce et al., 2020b). This raises questions about potential unintended or unorseen consequences, crucial enabling conditions (like legal changes or inter-institutional cooperation), as well as alignment with existing policy paradigms or advocacy coalitions (Howlett, 2020b). Evaluating through the lense of interconectedness and systemic effects helps to asses the "readiness for system change" identifying potential blockages to implementation that arise from conflicting interests or inertia in the public sector.

behavioural Insights, drawing from behavioural economics and psychology, call for a critical look at the assumptions built into the proposed solution about how citizens, front-line staff, or others would actually behave (Thaler & Sunstein, 2008; Olejniczak et al., 2020a). A policy design may appear reasonable on paper, but if it is implemented without recognising cognitive biases, decision making heuristics, or the strong impact of defaults and social norms, it might fail in practice (Shafir, 2013). Looking through a behavioural lens questions if the intervention is designed such that it recognises human psychology and organisational routines, and perhaps, recommend alterations such as simplification, nudges, or changes to the choice architecture that might facilitate uptake and effectiveness of the planned intervention (Datta & Mullainathan, 2014).

Futures Thinking clearly adds time-based considerations to the evaluation process, asking evaluators to account for resilience and adaptability of the proposed solution in the face of uncertain futures and evolving contexts (Conway et al., 2018; Government Office for Science, 2022). Futures Thinking involves determining the resilience of the initiative to different possible futures (i.e., economic changes, technological changes, or changes to social trends) and consider the long-term sustainability of the intervention (Schwartz, 1996; Wilkinson & Kupers, 2013). This way of considering options can help avoiding designing of initiatives that are aligned perfectly with the problems of today, but will quickly became either irrelevant or inefective as the world moves on, and can ultimately shift toward designs that incorporate flexibility and adaptability (Boyd & Osbourn, 2018).

The "Better Outcome" tool is intentionally designed to promote integrated perspectives within the evaluation criteria covering alignment with system level, behavoural relevance to the proposed change, and future readiness to deal with uncertainties context (See Section 1.4; See Chapter 4 for description). While fully incorporating these complex analyses within a rapid post-sprint evaluation is challenging, the tool serves as a crucial starting point for a discussion if such needemerges. It moves teams away from just evaluating the prototype artefact in isolation of each other and begins to consider interactions with the system, behavours, assumptions, and possible future's longevity.

In essence, applying these lenses converts any evaluation process to a version of anticipatory governance at a project level (Boyd & Osbourn, 2018). Traditional policy evaluations tend to look backward to assess impact after the solution has been implemented (Sanderson, 2002). Design sprints, however, deliver mostly early-stage concepts that demand an evaluation that looks forward instead. Integrating Systems Thinking, behavoural, and futures considerations means that the evaluation should be no longer focused just on judging a sprint process or on the the present prototype, but also anticipate future dynamics. It encourages teams to proactively identify potential complexities, behavoural pitfalls, and future vulnerabilities before committing substantial resources to the implementation. This anticipatory function is vital for de-risking innovation in the high-stakes environment of the public sector (Ahern, 2025; Tõnurist & Hanson, 2020). Ideally, it should invite a necessary discussion on viability and resilience that might otherwise be overlooked.

6.1.4. Practitioner Receptiveness: A Demand for Learning and Legitimacy

A particularly striking finding was the overwhelmingly positive attitude by practitioners towards an evaluation tool modelled in a structured format (Analysis 4.2.3). Survey responses indicated about 90% of participants believed a framework of this kind would be valuable (Finding 4.1.1) with interviewees and expert facilitators unanimously stating the same views (Findings 4.1.2, 4.1.3). This very high level of support goes against assumptions that people are engaged in creative, agile processes will dislike or resist a formal evaluation due to bureaucracy. Understanding the underlying reasons for this appreciation provides insight into the needs and aspirations of public sector innovators.

The demand for better evaluation mechanisms appears to be driven by a few things. First,

there is a prominent frustration with how things 'are', particularly the commonly noted implementation gap (Analysis 4.2.3). Practitioners felt disappointed when ideas, developed during intensive sprint efforts, did not live on (Finding 4.1.2). This promotes a desire for anything that would increase their chances of achieving follow-through and impact. They commonly viewed structured evaluation not as a barrier but as a potential vehicle for making ideas come to life.

Secondly, there is clearly a desire for learning and improvement. Interviewees indicated they wanted to know why some sprint outcomes thrive while others struggle, and they want to use that information to inform any future efforts (Finding 4.1.2, Analysis 4.2.3). A formal evaluation process provides an organised occasion for a facilitated reflection,feedback, and knowledge codification all in alignment with the principles of organisational learning in which they systematically reflect on experience to make future decisions (building upon concepts in Patton, 2015).

Thirdly, practitioners see a need for justification and legitimacy in their organisational context. Public sector environments typically require accountability for resource expenditures as well as evidence to support the decision-making (Newcomer et al, 2015). Interviewees referred to the value of having "something to show to my boss" (Finding 4.1.2), indicating that a formal evaluation output likely has a potential becoming an instrument of communicating the potential value and readiness of a sprint outcome, thereby supporting the case for future investment or strenghtening the political will. Thissuggests that even innovative processes must demonstrate their accountability within the parameters of established administrative frameworks.

Finally, the receptiveness is reinforced as the criteria by which the proposed tool was developed align closely with practitioners own internalised values, and dimensions for assessment (Analysis 4.2.2, 4.2.3). Since the evaluation tool formalises considerations that they already view as significant (for example, feasibility and public value), it appears relevant and supported rather than alien and imposed.

Taking these motivations into consideration holistically suggests that the evaluation tool offers practitioners both instrumental and symbolic value. Instrumentally, practitioners see it as an important tool to help improve decision making, inform learning, to rationalise cognitive load about resource use, and enhance the chances for successful implementation. Symbolically, using a formal evaluation framework may just as importantly enhance the

perceived legitimacy and professionalism of the Policy Design Sprint methodology itself, within the overall complex system of public administration. For public servants working in conditions where design-led processes may be perceived to be less evidence-based or systematic than traditional policy analysis (Section 1.2), the use of a formal evaluation framework may also be perceived as demonstrating a commitment to accountability, rigour and tangible outcomes, helping to bridge the cultural gap between experimental innovation and bureaucratic expectations. In this way, the tool is more than simply a checklist; it embodies the potential to further institutionalise and legitimise design-led practice in government.

6.2. Implications of the Research

The results and conclusions drawn above are important for both the theory and practice of public administration, policy design and innovation studies. This research adds to the academic discourse while also offering tangible tools and insights for practitioners seeking to enhance the effectiveness of design-led approaches in government.

6.2.1. Theoretical Contributions

The research makes several contributions to knowledge. First, it has filled a clear gap in the literature about the systematic assessment of Policy Design Sprint outcomes in the public sector (Section 1.1, Lit Review 2.1). The promise and process of sprints have been discussed (e.g., Bason, 2018; Kimbell, 2015b), but assessment of the outcomes of sprints, especially concerning policy feasibility and impact, has been largely absent from the literature. This research contributes by exploring and testing a distinct theory-based, multi-dimensional framework – the "Better Outcome" tool – designed to facilitate systematic assessment of Policy Design Sprint outcomes. This also shifts the focus away from easily measured,but shallow process metrics (e.g., the number of participants or ideas generated) towards assessing the quality and readiness of the outcomes (Section 1.2).

Secondly, the research adds value through its purposeful incorporation of sharing multiple disciplinary positions into one unified synthesising position. The framework drew insights from different areas, including Service Design (Stickdorn et al., 2018), Innovation Studies (Mulgan & Albury, 2003), Legal Design (Haapio & Hagan, 2020), Systems Thinking (Klein Woolthuis et al., 2005), Futures Thinking (Conway et al., 2018), and Behavioural Economics (Thaler & Sunstein, 2008). The multi-dimensional perspective provides a

means of understanding policy innovations in a more holistic and nuanced manner than any single disciplinary perspective would allow for (Sections 1.1, 1.3, 1.4, Lit Review 2.4). This interdisciplinary synthesis provides a conceptual model for how diverse theoretical streams can be operationalised to tackle complex evaluation challenges in applied settings.

Thirdly, the study provides valuable empirical insights into the dynamics of applying co-creative methodologies like Policy Design Sprints within the specific context of public administration. It provides an indication of the practical challenges encountered, in particular the ongoing implementation gap and the systemic factors contributing to it (Analysis 4.2.1). At the same time, it identifies opportunities, such as the tremendous tacit knowledge embedded in practitioners and their lasting desire to learn as well as "show value" (Analysis 4.2.2, 4.2.4). These findings contribute to a deeper understanding of the complex interplay between innovative processes and institutional contexts, particularly concerning the critical transition from collaborative ideation to tangible public impact, complementing existing work on co-creation and networked governance (Blomkamp, 2021a; Ansell & Torfing, 2014b).

6.2.2. Practical Contributions

Beyond its theoretical relevance, this research offers several direct practical contributions for those involved in public sector innovation. For Policy Design Sprint facilitators and participating teams, the study provides a concrete, validated instrument – the "Better Outcome" tool (Chapter 5) – designed to structure post-sprint reflection and assessment. This tool can help teams systematically evaluate the readiness of their generated concepts, identify potential risk and weaknesses early on, formalise their tacit knowledge, and improve communication about the sprint's results to stakeholders and decision-makers (Finding 4.1.2, 4.1.3, Analysis 4.2.2). It provides a shared language and framework for discussing outcome quality.

For public managers and public policymakers who are commissioning sprints or deciding what to do with the outputs of the sprints, the research provides a framework for gaining clearer and more structured insights on the potential value, feasibility and alignment of the ideas which emerged from the sprint (Finding 4.1.2). The evaluation framework provides a basis for comparing potential innovations against strategic priorities and making more informed decisions about resource allocation, piloting, or further development (Analysis

4.2.1, 4.2.2). It can help move decisions beyond subjective impressions towards evidence-informed choices.

For public sector innovation labs, policy labs, and similar units championing design-driven approaches, this work provides a means to systematically track, evaluate, and document the outcomes of their interventions. This capability is crucial for internal learning, iterative improvement of methods, and demonstrating value and impact to funders and political leadership (McGann et al., 2018). In an environment where such units sometimes face skepticism or budget pressures (Timeus & Gascó, 2018a; Wells, 2023; OECD 2023), the ability to systematically evaluate outcomes using a credible framework can strengthen their position and justify their continued operation (Section 1.2).

More generally, the research advances accountability and learning in the public organisations engaged in innovation. By advocating for and providing a tool for structured evaluation, it encourages a shift towards more evidence-based innovation practices (Patton, 2011). Furthermore, including reflection and an assessment component within the design process cultivates continuous improvement and supports learning from successes and failures through innovation. Thisultimately increases the likelihood that the design sprints meaningfully contribute to addressing public problems (Analysis 4.2.3, 4.2.4).

6.2.3. Implementing the "Better Outcome" Tool

The successful uptake and use of the “Better Outcome” evaluation tool is dependent not just on its design but also on how it is introduced and used as part of the Policy Design Sprint. Based on the findings of the research and feedback from practitioners and experts, following suggestions are offered.

Purpose and Framing: It is crucial to frame the tool's purpose correctly. First and foremost, it is a formative evaluation tool meant for the purpose of learning, improving, and measuring the readiness level of the sprint outcome for potential next steps (Finding 4.1.2). The tool should not be framed or seen as a summative judgement, a pass/fail test, or "grade" of the team. In other words, the "Better Outcome" tool is framed as a support tool to sustain structured conversation with the team, provide ground for collective sense making, and inform future action.

Timing of Use: The optimal time for use of the tool appears to be immediately at the conclusion of the Policy Design Sprint, possibly during the last afternoon or even during a

dedicated session not too long thereafter (Finding 4.1.2, 4.1.3).The timing is helpful in that participants should still have fresh memory of the context, discussions, and details of the constructed prototype. The expert feedback also pointed to the idea of framing the evaluation criteria right at the start of the sprint as being valuable, if only to orient the team, at the outset towards the characteristics of success from the future outcome and possibly influence the design process in its early phase (Finding 4.1.3).

Participants: Ideally the evaluation will involve the same core sprint team members who were engaged in the development of the outcome collaboratively. Their knowledge of the process and subsequent solution is valued for the real assessment to take place. Including a small number of key external stakeholders (the problem owner, implementers or represenative end-users) either in the evaluation session, or parts of it may be considered. External stakeholders can also provide broader perspectives, as well as reality checks. There is a careful consideration to assess in finding a balance between this inclusion, and the need for a psychologically safe environment for the core team to engage in a quite honest internal assessment without feeling too exposed or defensive. A professional judgment of the facilitator based on circumstance and group dynamics is needed for making such decision.

Facilitation: The process of evaluation needs to be facilitated so that the team is progressing through the "Better Outcome" tool assessment design (detailed in Chapter 5). The facilitator needs to guide the team through each of the criteria, promoting discussion, encouraging sharing of evidence or rationale behind ratings, and documenting key findings, any assumptions surfacing, critical risk and next (possible/ideas) steps. This is about constructive dialogue and creating shared understanding rather than simply tagging marks.

Understanding Results: Participants should be encouraged to see the evaluation output as more than a total score or judgement, but rather as a useful profile of the strengths, weaknesses, and readiness of a sprint overall. The actual scores for each dimension of evaluation are less important than the map of the ratings that define the profile. The profile is meant to assist in identifying and acting upon critical risks or weaknesses (i.e. poor feasibility or stakeholder buy-in, seriousmisalignment) that need immediate attention or need to be further appraised before the idea can realistically progress.

Transitioning to Next Steps: The evaluation should not be considered in isolation. The output of the evaluation is intended to direct next actions and decisions regarding the sprint output. Both the team and decision-makers should follow the evaluation profile to: move to piloting or testing; revise the concept based on areas of weakness; be political to seek approvals or resources; pivot the idea, if in a meaningful way; or even conclude that nothing further is worthwhile if the there are fundamental errors or deal-breakers. The evaluation should end with a clear action plan or recommendation that indicated ownership and responsibility, regardless of whether there are actions to move forward or to terminate.

6.2.4. Potential Challenges and Mitigation

Practitioners, as well as experts, identified several potential challenges in conducting and evaluation:

Time pressure: The time commitments for sprints are considerable, and evaluation will take time. Mitigation involves developing the tool to be both short and easy to use; conducting the session efficiently; and explicitly allocating enough time for evaluation either as part of the sprint agenda or immediately following (Finding 4.1.2, 4.1.3).

Subjectivity and Bias: The team is likely to have a natural attachment to what they have created and be overly-optimisticin their evaluations (confirmation bias). Mitigation includes motivating the teams to anchor their ratings to whatever forms of evidence they have (even if feedback from users was limited in terms of being preliminary, or they made some rough feasibility checks), encouraging critical reflection during the facilitation, possibly including an outside facilitator or stakeholder perspectives to help inject processes for objectivity, and being clear about assumptions (Finding 4.1.2).

Lack of Data: The outcome of a sprint is likely to be an early prototype without sufficient empirical data available for validation. Mitigation includes framing the evaluation ratings as based on current knowledge and highlighted identified assumptions. The tool itself has the potential to be useful in identifying the data gaps and thus identify the uncertainty, thus providing a priority for future research, or testing, or validation of activities.

Organisational Culture: The use of the tool can be significantly undermined in organisational culture that is risk-averse, assigns blame when failure looks like it will happen, or lack genuine support for innovation and does not authentically support innovation. Mitigations will require addressing organisational culture through framing the

purpose of the tool (learning, not judgement), getting explicit buy-in from leadership that evaluation is intended to be constructive and supportive, checking that psychological safety is maintained through the evaluation session, and showing that what was learned will be used for improvement, not punishment (Analysis 4.2.3).

Ensuring Follow Through: The evaluation tool provides a sense of readiness and expectations but cannot by itselfguarantee an implementation. Mitigation will require actively linking the evaluation output with any existing organisational decision making processes, making sure to identify a project champion or owner to be responsible for further steps, making pathways clear for obtaining the resources or approvals based on the evaluation outcome (Finding 4.1.1, 4.1.2, Analysis 4.2.1).

By considering these limitations and implications in the planning Policy Design Sprint project, organisations will increase the chances that the "Better Outcome" evaluation tool will be used effectively, and that it will contribute to the impact and sustainability of the Policy Design Sprint process.

6.3. Limitations and Future Research Directions

This research offers useful ideas and a tool for evaluating outcomes from a Policy Design Sprint, but there are identified limitations to this study which provide space for future research. Recognising these limitations places the findings in context and will help shape efforts undertaken to build on this first study.

6.3.1. Methodological Limitations

The study primarily employed a qualitative, exploratory research strategy (Section 3.1). While this approach yielded rich, contextual insights into the experiences and perspectives of practitioners in Latvia, the findings are based on a relatively small number of participants (21 survey respondents, 4 interviewees, 3 experts) selected purposively (Section 3.2, 3.3). Consequently, the results may lack statistical generalizability to the broader population of public sector innovators or sprint activities (Section 1.4). The reliance on self-reported data through surveys and interviews also introduces the potential for recall bias or social desirability bias, although triangulation across methods aimed to mitigate this.

A significant limitation is that the "Better Outcome" evaluation tool, while developed

iteratively and validated through expert walkthroughs, was not tested in a live Policy Design Sprint setting as part of this research project (Section 3.3). Therefore, its real-world usability, the time required for its application, its influence on team dynamics, and its actual impact on decision-making and implementation success remain to be empirically verified through field testing.

Furthermore, the research scope was deliberately focused on the immediate post-sprint evaluation of outcomes (Section 1.4). It did not assess the quality of the sprint facilitation process itself, nor did it track the long-term trajectory and ultimate impact of the sprint-generated ideas months or years after the initial evaluation. Understanding the full lifecycle of these innovations requires longitudinal research designs.

6.3.2. Contextual Limitations

The empirical data for this study were collected exclusively within the context of the Latvian public sector (Section 1.4). While the challenges identified (e.g., implementation gap) and the criteria deemed important by practitioners resonate with international literature, the specific institutional culture, political environment, and administrative traditions of Latviaundoubtedly shape the local experience of Policy Design Sprints. Therefore, the direct transferability of the findings and the applicability of the "Better Outcome" tool in regional, or organisational contexts should be approached with caution and may require adaptation (Section 1.4).

6.3.3 Future Research Directions

The limitations discussed in previous subsections point toward some promising future research avenues:

Quantitative Validation: Quantitative studies incorporating larger scale surveys in public sector organisations and possibly in different countries could help to quantify the prevalence of the implementation gap, identify any statistically significant factors impacting sprint success and test the generalizability of the evaluation criteria identified in this study.

Live Piloting and Refinement: the next most significant step is to pilot the "Better Outcome" tool across a number oflive Policy Design Sprints in a various policy domains and organisational contexts. This would include using it to observe its use, obtain feedback

from participating teams and facilitators on its usability and perceived value, and a subsequent iterative process of refining the design of the tool, weighting of the criteria and guidance on its use based on real-life usage.

Longitudinal Impact Studies: Future research should use longitudinal designs to follow sprint outcomes beyond simplypost-sprint evaluations. This could be used to get a better appreciation on the long-term impact of the ideas implemented, the elements that predict longer successes vs recent adopt and abandon behaviour, and potentially evaluate the predictive validity of the "Better Outcome" tool initial evaluation.

Comparative Studies: Comparative research designs could help examine the effectiveness of sprints that are using a structured evaluation tool vs sprints that are not, potentially using quasi-experimental strategies where possible to determine the differences in effect. Comparing the use of the tool and utility of the evaluation criteria across types of policy problems (e.g. service delivery vs. disaster response) may also yield useful data. Regulatory formats, or organisational cultures, could also produce relevant responses to further comparative studies.

Inquiring into Tool Adaptation: Research could address how the core principles and criteria of the "Better Outcome" tool would be adapted, or tailored, for assessing outcomes of different types of design-led interventions in the public sector, particularly shorter co-creation workshops or longer-term co-design projects. The adaptations could also be considered for specific policy domains (digital transformation, healthcare, environmental policy).

Investigating Organisational Integration: Another important area for future research is how public organisations can effectively embed structured evaluation mechanisms for design-led innovations into their established governance structures and processes (e.g. service delivery, budget cycles, performance management frameworks, risk management). This would include organisational importance of organisational culture change and leadership commitment, as well as capacity building in by adopting and utiliseing tools like the "Better Outcome" framework and ensuring evaluation leads to meaningful action.

Exploring these avenues of research will help to further inform how to pursue design approaches to maximise value and impact within the complexities of public administration. For example, to step away from the limitations and resulting promising research directions,

this research opens the opportunity for a wider discussion on other aspects of Design Policy Sprints. The outcomes are mostly direct answers to the problem that was brought to the team for solving through the sprint by employing the agile innovative methodologies. The nature of the problem is not always suitable for design sprint as a format therefore outcome is forced through unsuitable methodology and approach. Disappointment is always the result of unfulfilled expectations. The evaluation of the suitability of methodology for a particular problem might be as important as evaluation of the outcome, which indicates another promising field for research which is not directly addressed by this thesis.

7. CONCLUSION

In summary, this research addresses a critical gap in public-sector innovation: how to evaluate the outcomes of Policy Design Sprints in government contexts. The study was motivated by the rapid adoption of design-sprint methods in public policymaking—especially in the Latvian government—in response to complex challenges (e.g., climate change, pandemics) that demand innovative, collaborative solutions. Drawing on Service Design, Innovation Studies, Legal Design, Systems Thinking, behavioural Economics, and methods of Futures Thinking, the study proposed and tested a structured evaluation framework (the "Better Outcome" tool) to bridge theory and practice in assessing sprint-generated policy ideas. A qualitative, interpretivist–constructivist methodology was employed, involving surveys of sprint participants, semi-structured interviews with experts, and iterative design-prototype development. This approach included multiple stakeholder perspectives and recognized that policy reality is socially constructed by practitioners (Crotty, 1998; Creswell, 2013). Data from practitioners' experiences with policy sprints were triangulated with expert feedback, and the iterative design process integrated these insights into a usable assessment tool. In line with the interpretivist stance, the researcher acknowledged that meaning emerges through interaction with participants and maintained reflexivity about personal assumptions during data collection and analysis.

The central findings of the study illuminate both the promises and persisting challenges of Policy Design Sprints in the public sector, and they validate the potential of the “Better Outcome” tool to address those challenges. The key findings are summarized below:

Persistent implementation gap: Public-sector design sprints tend to generate creative policy concepts, but very few of these concepts are turned into practice. Participants reported that ideas "often stall once the sprint ends" due to multiple factors such as unclear ownership, lack of political support, institutional misfit or existance of other feasibility barriers. This aligns with the literature on policy innovation that identifies a "design–implementation gap" (Ansell & Torfing, 2014b; van Buuren et al., 2020a) and underscores the need for follow-up mechanisms. In practical terms, sprint teams frequently expressed anxiety that without deliberate next steps, their outputs would never be implemented. As one participant noted, "We ended with a great prototype, but no one

was sure who should carry it forward." This finding highlights the importance of incorporating accountability and transition strategies into the sprint process—ensuring that "what gets measured, gets done" (Osborne & Gaebler, 1992) and that high-potential innovations do not languish unrealized.

Multi-dimensional success criteria: Sprint participants and facilitators defined "success" far more broadly than a simple prototype or report. Success included practical feasibility (budgetary, legal, organizational feasibility), stakeholder buy-in (acceptance by decision-makers and affected citizens), substantive public value (the extent to which the solution meaningfully addresses the problem), sustainability over time, and systemic fit. These practitioner-generated criteria closely matched the evaluative dimensions advocated in the academic literature (e.g., Kimbell, 2016b; Tõnurist et al., 2017; De Vries, Bekkers, & Tummers, 2016). In effect, the study found that "the team's checklist and the academic checklist turned out to be one and the same." This consensus validates the approach of adopting Systems Thinking and a Futures perspective into evaluation, making it possible to incorporate public innovators' values and concerns into future evaluation. The study reinforced that evaluation tools must address this multifaceted nature, and evaluate a concept's originality, viability, acceptability and systemic effect.

Enthusiasm for structured evaluation: Contrary to the stereotype that creative teams resist structure, the vast majority of practitioners welcomed the idea of a dedicated outcome assessment. In the survey, roughly 90% of respondents agreed that a structured evaluation framework would benefit their projects, and all interviewees and expert reviewers supported the concept. Rather than seeing evaluation as bureaucratic red tape, participants viewed it as a practical aid and learning opportunity. As one Policy Design Sprint facilitator remarked, "We need this kind of reflection – it helps us justify our work and improve next time." This strong bottom-up support indicates that the field of public innovation is ready for more evidence-based practice. Practitioners want to know not just that they "did a sprint well," but how to ensure that the sprint's output is more likely to succeed post-sprint (De Vries et al., 2016). Embedding evaluation within the sprint process was viewed instrumentally (to improve outcomes) and symbolically (to lend legitimacy to design sprints within bureaucratic contexts).

Validation of the Better Outcome tool: The proposed Better Outcome evaluation framework was established as a viable solution through testing with experts. Subject matter

experts who walked through the tool with the researcher agreed the dimensions of the tool were comprehensive and that the tool was parctical and fit for use. They particularly liked the intuitive Likert-scale questions and the radar-chart visualization that provides immediate feedback which can be used also for further discussion with decision makers. Experts suggested only minor refinements—such as clarifying some wording and adding a prompt to explicitly assign responsibility for follow-up—to optimize clarity and completeness. These changes have been made, and the prototype is ready for use and further development. Most importantly, the experts confirmed that the criteria and process built into the tool fit well with practitioners' values and workflow; they viewed the "Better Outcome" tool as a natural fit into sprints rather than an interruption to work they were already doing. In sum, the "Better Outcome" framework was judged ready for field use: easy to use, in line with known good practices, and likely to facilitate better decision-making, knowledge transfer, and next steps after the sprint. The framework thus appears poised to help close the implementation gap by making follow-up and viability a transparent and shared concern.

Together, these findings paint a coherent picture: policy teams have been mainly conducting design sprints "in the dark," without a clear means to assess or ensure impact. The Better Outcome tool intervenes by **bringing visibility to outcomes**. By structuring reflection around core criteria and providing visual benchmarks, the tool helps teams and managers answer questions such as "Are we solving the right problem?", "How feasible and acceptable is our solution?",and "What still needs attention and input?". In the words of an expert reviewer, the radar chart "turns a fuzzy conversation into an actionable discussion." This is consistent with the broader goal of evaluation in government: to provide credible, useful information that enables lessons learned to be incorporated into decision-making. The tool thus aligns with OECD definitions of evaluation as a systematic and objective assessment of policy initiatives and with the aspiration that evaluation should improve plans and practices for citizens. In practical terms, the use of the tool during or immediately after a sprint would help transform implicit team insights into explicit data, thereby fostering organizational learning and accountability.

Contributions to Scholarship and Practice. This research makes several original contributions. The proposed evaluation framework might also be viewed as interdisciplinary with an appreciation for different schools of thought. Although it does not

claim to solve the competing paradigmatic perspective challenges from Service Design, public innovation theory, legal and futures perspectives, and behavioural insights together, it may be more comprehensive than other articulations of holistic assessments of policy innovation. This integration is novel: it operationalizes criteria from domains that rarely speak directly to each other (e.g. Futures Thinking and Legal Design) into a unified tool. In doing so,the research provides a conceptual model for how such diverse criteria can be operationalized in practice. Empirically, the study fills a gap in the literature by documenting how design sprints are actually perceived and implemented within public administration. It documents the lived experience of sprint teams and the specific barriers they face. In particular, it confirms that Service Design approaches, while promising in principle, often clash with bureaucratic norms and lack of follow-through (as noted in prior work by Ansell & Torfing, 2014b; Kimbell, 2016b). At the same time, it uncovers the generally untapped capacity of public servants to evaluate and learn from innovation projects, extending evidence of a field-wide shift toward reflective practice (De Vries et al., 2016).

Practically, the core output is the “Better Outcome” tool itself (including the evaluation questionnaire, radar-graph visualization, and guidelines for use). This is a direct design contribution: a tangible resource that public-sector innovators can apply to improve their sprints. It codifies the collective wisdom of experts and past research into a simple format. The toolkit can help move policy sprints from an experimental novelty to a more accountable innovation process by explicitly prompting teams to address issues like ownership of ideas, stakeholder buy-in, and ethical considerations. It may also provide organizations with a benchmarking tool: for example, civil servants could compare their sprint outcomes to estimations or target values, or across policy domains, to potentially support shared learning and encourage healthy competition. In this way, the research aids practitioners in demonstrating "public value" and building trust in design-led policymaking rather than relying on "superficial process metrics" like the number of participants or outputs.

Researcher Reflexivity and Interpretivist Stance. Throughout the study, the researcher remained mindful of the interpretivist–constructivist paradigm underpinning the work. This stance acknowledges that knowledge is co-created: the criteria and insights generated in the "Better Outcome" tool emerge from the meanings that practitioners attach to their

experiences. The researcher's role was thus as a facilitator and synthesizer of those meanings, rather than an "objective" observer. In practice, this meant using open-ended interviews and "walkthrough workshops" to elicit how participants *interpret* success in their own work, and then iteratively refining the tool in partnership with them. Reflexivity was maintained by constant dialogue with participants (e.g., sharing intermediate findings for feedback) and by transparency about the study's purpose. The questions were informed by the researcher's background in Service Design and legal science. This self-awareness supports best practice for qualitative inquiry (Lincoln & Guba, 1985) and provides confidence that participants' perspectives are reflected in the conclusions.

Limitations and Future Directions. This study has several limitations that point to avenues for future research. The most significant limitation is the lack of field deployment of the “Better Outcome” tool in a live Policy Design Sprint. The present evaluation of the tool was based on hypothetical scenarios and expert walkthroughs, which are invaluable for formative testing but cannot capture all the dynamics of a real sprint. Future research should, therefore, pilot the tool in actual government sprints across various policy domains and levels of government. Such field trials would reveal practical issues (e.g., time required to complete the assessment, group facilitation effects, technological requirements if digitized) and allow measurement of downstream impact (e.g., whether scores on the tool predict actual implementation success or service outcomes).

Another limitation is the study's geographic and organizational scope. Data were gathered exclusively from Latvian public-sector innovators and experts. While many findings likely extrapolate to similar contexts, cultural and institutional differences (e.g., in bureaucracy, political oversight, or design maturity) could influence outcomes. A comparative research program across countries and agencies would be relevant to test the tool's applicability to other contexts to determine context-specific elements. The sample of surveys and interviews, while representative across agencies, was modest. More robust empirical studies (potentially quantitative, multi-site) could strengthen the evidence.

Several open questions also emerged about the broader practice of Policy Design Sprints. For example, some participants noted that not every policy problem may be a good fit for a sprint format. Systematically evaluating when to use a sprint versus other methods – perhaps by adding a "fit-for-purpose" assessment in the toolkit – could be an important extension. Moreover, further inquiries could take the organizational culture into account as

a mediating factor in the uptake of evaluation in the practice, as this culture may heavily influence practice.

Perspectively, this research may inform future evaluations of Policy Design Sprint outcomes in the public service. As agile innovation methods gain traction, so does the expectation for accountability and learning. The "Better Outcome" tool exemplifies how this trend can be institutionalized by making evaluation a standard sprint deliverable. Over time, data accumulated from using such tools could feed into general evaluations of public innovation, helping to identify which sprint practices tend to yield real change. Moreover, integrating evaluation may create the conditions for engaging career public servants who are sceptical about the worth of agile methodologies by showing tangible evidenced of the generated value. From a broader view, this research means that research and practice continue to nudge the public service toward evidence informed co-creation public administration process.

Answering the central research question directly - the research concludes that evaluating the outcomes of Policy Design Sprints should be done using a defined multi-dimensional framework (such as the proposed "Better Outcome" tool) that assesses the outcome according to the criteria of: implementation viability, systemic relevance, behavioural insights, future opportunities to adapt etc., thus giving policy-makers a complete evidence-supported assessment of practical opportunity and strategic significance.

To conclude, this inquiry closes the loop between the theory and practice of service-design-led policy innovation, showing that thoughtful evaluation can co-exist with any kind of creative process and that designers and policymakers share much more than is sometimes assumed. By foregrounding Systems Thinking, stakeholder perspectives, and actionable knowledge transfer, the “Better Outcome” framework offers a practical way to make design sprints more than just an exercise in creativity. The significance of this research lies in its potential to make government innovation not only more inventive but also more effective and accountable. Although challenges remain, the path forward is clear: systematic evaluation must become a routine part of the Policy Design Sprint, so that promising ideas can translate into the “better outcomes” they were meant to achieve.

LIST OF REFERENCES

Ahern, D. (2025). The new anticipatory governance culture for innovation: Regulatory foresight, regulatory experimentation and regulatory learning. European Business Organization Law Review, 26(1), 1–33. https://doi.org/10.1007/s40804-025-00348-7

Ansell, C., & Torfing, J. (2014a). Public innovation through metagovernance and collaborative governance. In S. P. Osborne & L. Brown (Eds.), Handbook of innovation in public services (pp. 467–483). Edward Elgar Publishing.

Ansell, C., & Torfing, J. (Eds.). (2014b). Public innovation through collaboration and design (1st ed.). Routledge.

Bason, C. (2017). Leading public design: How managers can engage citizens in co-production and co-creation. Policy Press.

Bason, C. (2018). Leading public sector innovation: Co-creating for a better society (2nd ed.). Policy Press.

Bason, C., & Austin, R. D. (2022). Design in the public sector: Toward a human-centred model of public governance. Public Management Review, 24(11), 1727–1757. https://doi.org/10.1080/14719037.2021.1919186

Bernard, H. R. (1988). Research methods in cultural anthropology. Sage Publications.

Blomkamp, E. (2018). The promise of co-design for public policy. Australian Journal of Public Administration, 77(4), 729–743. https://doi.org/10.1111/1467-8500.12310

Blomkamp, E. (2021a). Co-creation and public innovation: A comparative analysis of methods, types, and aims. Policy and Society, 40(4), 516–533. https://doi.org/10.1080/14494035.2021.1980864

Blomkamp, E. (2021b). Systemic design practice for participatory policymaking. Policy Design and Practice, 5(1), 12–31. https://doi.org/10.1080/25741292.2021.1887576

Borins, S. (2014). The persistence of innovation in government: A guide for innovative public servants. IBM Center for the Business of Government. https://www.businessofgovernment.org/report/persistence-innovation-government-guide-innovative-public-servants

Boyd, A., & Osbourn, G. (2018). Anticipatory governance: Improving foresight capability in government. Nesta.

https://media.nesta.org.uk/documents/Anticipatory_Governance_Improving_foresight_capability_in_government.pdf

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, T. (2009). Change by design: How design thinking transforms organizations and inspires innovation. HarperBusiness.

Bryson, J. M., Crosby, B. C., & Bloomberg, L. (2014). Public value governance: Moving beyond traditional public administration and the new public management. Public Administration Review, 74(4), 445-456. https://doi.org/10.1111/puar.12238

Cambridge Assessment. (2023). The futures of assessment: Navigating uncertainties through the lenses of anticipatory thinking. https://www.cambridgeassessment.org.uk/Images/698413-the-futures-of-assessment-navigating-uncertainties-through-the-lenses-of-anticipatory-thinking.pdf

Chan, F. K., Thong, J. Y., Brown, S. A., & Venkatesh, V. (2025). Design characteristics and service experience with e-government services: A public value perspective. International Journal of Information Management, 80, 102834. https://doi.org/10.1016/j.ijinfomgt.2024.102834

Christiansen, J. (2014). Innovation labs: Leveraging openness for public sector innovation? In S. P. Osborne & L. Brown (Eds.), Handbook of innovation in public services (pp. 289–300). Edward Elgar Publishing.

Clarke, A., & Craft, J. (2019). The twin faces of public sector design. Governance, 32(1), 5–21. https://doi.org/10.1111/gove.12342

Cohen, D., & Crabtree, B. (2006). Qualitative research guidelines project. Robert Wood Johnson Foundation. http://www.qualres.org/HomeEval-3664.html

Conway, M., Masters, J., & Thorold, J. (2018). Strategic foresight in government: A review of state-of-the-art practices. Policy Lab UK. https://openpolicy.blog.gov.uk/wp-content/uploads/sites/161/2018/07/Strategic-Foresight-in-Government-report-Policy-Lab-July-2018.pdf

Creswell, J. W. (2013). Qualitative inquiry & research design: Choosing among five approaches (3rd ed.). Sage Publications.

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). SAGE.

Crotty, M. (1998). The foundations of social research: Meaning and perspective in the research process. Sage Publications.

Datta, S., & Mullainathan, S. (2014). Behavioral design: A new approach to development policy. Review of Income and Wealth, 60(1), 7–35. https://doi.org/10.1111/roiw.12109

De Vries, H., Bekkers, V., & Tummers, L. (2016). Innovation in the public sector: A systematic review and future research agenda. Public Administration, 94(1), 146–166. https://doi.org/10.1111/padm.12209

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2018). The SAGE handbook of qualitative research (5th ed.). Sage Publications.

Dorst, K. (2011). The core of "design thinking" and its application to policy innovation. Design Studies, 32(6), 521–532. https://doi.org/10.1016/j.destud.2011.07.006

Dow, S. P., Fortuna, J., Schwartz, D., Altringer, B., & Klemmer, S. R. (2012). Prototyping dynamics: Sharing multiple designs improves exploration, group rapport, and design outcomes. ACM Transactions on Computer-Human Interaction (TOCHI), 19(2), Article 11. https://doi.org/10.1145/2240156.2240162

Dufva, M. (2019). Using the megatrends. Sitra. https://www.sitra.fi/en/articles/using-the-megatrends/

Dunn, W. N. (2017). Public policy analysis (6th ed.). Routledge.

Edquist, C. (2011). Design of innovation policy through diagnostic analysis: Identification of systemic problems (or failures). Industrial and Corporate Change, 20(6), 1725–1753. https://doi.org/10.1093/icc/dtr060

Edquist, C. (2020). Public procurement for innovation (PPI): A research agenda. Innovation: The European Journal of Social Science Research, 33(3), 269–283. https://doi.org/10.1080/13511610.2020.1757114

Fowler, F. J., Jr. (2014). Survey research methods (5th ed.). SAGE Publications.

Gailīte, D. (2023, January 17). Metodika likumprojekta kvalitātes mērīšanai: Gunāra Kusiņa izstrādātais pārbaudes tests. [Methodology for evaluating quality of legal drafts: A test developed by Gunārs Kusiņš] Jurista Vārds, (3), 10–11. https://juristavards.lv/doc/282621-metodika-likumprojekta-kvalitates-merisanai-gunara-kusina-izstradatais-parbaudes-tests/

Government Office for Science. (2022). The Futures Toolkit: Tools for futures thinking across UK government. https://assets.publishing.service.gov.uk/media/62387612d3bf7f4f1b1de497/futures-toolkit-march-2022.pdf

Government Office for Science. (2024). Futures toolkit for policy makers and analysts. https://www.gov.uk/government/publications/futures-toolkit-for-policy-makers-and-analysts/the-futures-toolkit-html

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. Field Methods, 18(1), 59-82. https://doi.org/10.1177/1525822X05279903

Gustetic, J., Teixeira, C., Carroll, B., Cheung, J., O'Malley, S., & Brewster, M. (2020, May 6). Policy prototyping for the future of work. CoLab Harvard Kennedy School. https://ash.harvard.edu/wp-content/uploads/2024/02/colab-hks_5-6-2020_1_1.pdf

Haapio, H., & Hagan, M. (2020). Design patterns for contracts: Combining legal design with computational thinking. Proceedings of the 2020 ACM Conference on Human Factors in Computing Systems (CHI '20) (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376148

Haapio, H., & Hagan, M. (Eds.). (2021). Legal design: Integrating business, design and legal thinking with technology. Edward Elgar Publishing.

Haynes, L., Service, O., Goldacre, B., & Torgerson, D. (2012). Test, learn, adapt: Developing public policy with randomised controlled trials. Cabinet Office Behavioural Insights Team. https://www.bi.team/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials/

Hermus, M., Van Buuren, A., & Bekkers, V. (2020). Applying design in public administration: A literature review to explore the state of the art. International Review of Administrative Sciences, 86(3), 411-430. https://doi.org/10.1177/0020852319841191

Holmlid, S., & Evenson, S. (2008). Bringing Service Design to service sciences, management and engineering. In B. Hefley & W. Murphy (Eds.), Service science, management and engineering: Education for the 21st century (pp. 341-345). Springer. https://doi.org/10.1007/978-0-387-76578-5_49

Howlett, M., Ramesh, M., & Perl, A. (2009). Studying public policy: Policy cycles and policy subsystems (3rd ed.). Oxford University Press.

Howlett, M. (2018). Designing public policies: Principles and instruments (2nd ed.). Routledge.

Howlett, M. (2020a). Policy design: Towards a new generation of analysis tools and skills. Políticas Públicas & Cidades, 8, 1-17. https://portalseer.ufba.br/index.php/ppc/article/view/35638

Howlett, M. (2020b). Policy paradigms: Choosing perspectives on public policy. In M. Moran, M. Rein, & R. E. Goodin (Eds.), The Oxford Handbook of Public Policy. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198829705.013.2

Howlett, M. (2023). Designing public policies: Principles and instruments. Routledge.

Huić, A., Horvat, A., & Škec, S. (2023). Design sprint: Use of design methods and technologies. Proceedings of the Design Society, 3, 2645-2654. https://doi.org/10.1017/pds.2023.265

Jones, P., & Bowes, J. (2017). Rendering systems visible for design: Synthesis maps as constructivist design narratives. She Ji: The Journal of Design, Economics, and Innovation, 3(3), 229–248. https://doi.org/10.1016/j.sheji.2017.12.001

Kimbell, L. (2015a). Applying design approaches to policy making: Discovering Policy Lab. University of Brighton. https://ualresearchonline.arts.ac.uk/id/eprint/8851/

Kimbell, L. (2015b). The service innovation handbook: Action-oriented creative thinking tools and methods. BIS Publishers.

Kimbell, L. (2016a). Rethinking capacity building: Organisational change and design practices. Design Issues, 32(4), 113–126. https://doi.org/10.1162/DESI_a_00430

Kimbell, L. (2016b). Rethinking the boundaries of public Service Design. Information Polity, 21(4), 403–417. https://doi.org/10.3233/IP-160397

Kimbell, L., & Bailey, J. (2017). Prototyping and the new spirit of policy-making. CoDesign, 13(3), 214–226. https://doi.org/10.1080/16108167.2017.1355003

Klein Woolthuis, R., 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

Knapp, J., Zeratsky, J., & Kowitz, B. (2016). Sprint: How to solve big problems and test new ideas in just five days. Simon & Schuster.

Knetsch, J. (2011). Behavioural economics, policy analysis and the design of regulatory reform. In D. Low (Ed.), Behavioural economics and policy design: Examples from Singapore (pp. 161–182). World Scientific. https://doi.org/10.1142/9789814340897_0007

Koskinen, I., Zimmerman, J., Binder, T., Redstrom, J., & Wensveen, S. (2011). Design research through practice: From the lab, field, and showroom. Morgan Kaufmann.

Lewis, J. M., McGann, M., & Blomkamp, E. (2020). When design meets power: Design thinking, public sector innovation and the politics of policymaking. Policy & Politics, 48(1), 111–130. https://doi.org/10.1332/030557319X15613695781113

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Sage Publications.

Manzini, E. (2015). Design, when everybody designs: An introduction to design for social innovation. The MIT Press.

McGann, M., Blomkamp, E., & Lewis, J. M. (2018). The rise of public sector innovation labs: Experiments in design thinking for policy. Policy Sciences, 51(3), 249–267. https://doi.org/10.1007/s11077-018-9315-7

McGann, M., Wells, T., & Blomkamp, E. (2021). Innovation labs and co-production in public problem solving. Public Management Review, 23(2), 297-316. https://doi.org/10.1080/14719037.2019.1699946

Mercure, J.-F., Lam, A., Egging-Bratseth, R., & Pollitt, H. (2021a). Modelling innovation and the macroeconomics of low-carbon transitions. Research Policy, 50(10), 104368. https://doi.org/10.1016/j.respol.2021.104368

Mercure, J.-F., Salas, P., Vercoulen, P., Semieniuk, G., Ahmad, S., Pollitt, H., Holden, P. B., Vakilifard, N., Chewpreecha, U., Edwards, N. R., & Vinuales, J. E. (2021b). Reframing incentives for climate policy action. Nature Energy, 6, 1133–1143. https://doi.org/10.1038/s41560-021-00934-2

Mergel, I., Gong, Y., & Bertot, J. (2018). Agile government: Systematic literature review and future research. Government Information Quarterly, 35(2), 291-298. https://doi.org/10.1016/j.giq.2018.04.003

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). Qualitative data analysis: A methods sourcebook (3rd ed.). Sage Publications.

Mintrom, M., & Luetjens, J. (2016a). Design thinking in policymaking processes: Opportunities and challenges. Australian Journal of Public Administration, 75(3), 391–402. https://doi.org/10.1111/1467-8500.12211

Mintrom, M., & Luetjens, J. (2016b). Policy entrepreneurs and the diffusion of innovation: A systematic review of the literature. Policy Studies Journal, 44(S1), S110-S134. https://doi.org/10.1111/psj.12159

Monteiro, B., & Kumpf, B. (2023, December 18). Innovation labs through the looking glass: Experiences across the globe. OECD OPSI. https://oecd-opsi.org/blog/innovation-labs-through-the-looking-glass/

Moore, M. H. (1995). Creating public value: Strategic management in government. Harvard University Press.

Mulgan, G., & Albury, D. (2003). Innovation in the public sector. Strategy Unit, Cabinet Office. https://webarchive.nationalarchives.gov.uk/ukgwa/20100407172811/http://www.cabinetoffice.gov.uk/media/cabinetoffice/strategy/assets/innovation.pdf

Neumann, O., Kirklies, T., & Schott, C. (2024). Adopting agile in government: A comparative case study. Public Management Review, 26(12), 3692-3714. https://doi.org/10.1080/14719037.2024.2335109

Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (Eds.). (2015). Handbook of practical program evaluation (4th ed.). Jossey-Bass

Nielsen, J. (1993). Usability engineering. Academic Press.

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), Usability inspection methods (pp. 25–62). John Wiley & Sons.

Nogueira, C. A., & Schmidt, S. (2022). Innovation for better management: The contribution of public innovation labs. Inter-American Development Bank. https://doi.org/10.18235/0004067

Norman, D. A. (2013). The design of everyday things: Revised and expanded edition. Basic Books.

OECD Observatory of Public Sector Innovation. (2023). Futures and foresight. OECD OPSI. https://oecd-opsi.org/guide/futures-and-foresight/

Olejniczak, K., Śliwowski, P., & Leeuw, F. L. (2020a). Behavioral insights for public policy: A systematic literature review. European Policy Analysis, 6(1), 116–142. https://doi.org/10.1002/epa2.1082

Olejniczak, K., Śliwowski, P., & Leeuw, F. L. (2020b). Comparing behavioral assumptions of policy tools: Framework for policy designers. Journal of Comparative Policy Analysis: Research and Practice, 22(6), 498-520. https://doi.org/10.1080/13876988.2020.1808465

Organisation for Economic Co-operation and Development. (2019). Embracing innovation in government: Global trends 2019. OECD Publishing. https://doi.org/10.1787/9789264310707-en

Organisation for Economic Co-operation and Development. (2020). Systemic thinking for policy making: The potential of systems analysis for addressing global policy challenges in the 21st century. OECD Publishing. https://doi.org/10.1787/879c4f7a-en

Organisation for Economic Co-operation and Development. (2021). The design and implementation of mission-oriented innovation policies: A new systemic policy approach to address societal challenges (OECD Science, Technology and Industry Policy Papers, No. 100). OECD Publishing. https://doi.org/10.1787/3f6c76a4-en

Organisation for Economic Co-operation and Development. (2023). Public sector innovation labs: Stocktaking and future prospects. OECD Publishing. https://doi.org/10.1787/953939d7-en

Osborne, D., & Gaebler, T. (1992). Reinventing government: How the entrepreneurial spirit is transforming the public sector. Addison-Wesley.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Administration and Policy in Mental Health and Mental Health Services Research, 42(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y

Parker, S., & Heapy, J. (2006). The journey to the interface: How public Service Design can connect users to reform. Demos. https://www.demos.co.uk/files/Journey%20to%20the%20Interface%20-%20web.pdf

Patton, M. Q. (2011). Developmental evaluation: Applying complexity concepts to enhance innovation and use. Guilford Press.

Patton, M. Q. (2015). Qualitative research & evaluation methods: Integrating theory and practice (4th ed.). Sage Publications.

Pawson, R., & Tilley, N. (1997). Realistic evaluation. Sage Publications.

Peters, B. G. (2018a). Advanced introduction to public policy (2nd ed.). Edward Elgar Publishing.

Peters, B. G. (2018b). The politics of bureaucracy: An introduction to comparative public administration (7th ed.). Routledge.

Pierce, J. J., Peterson, H. L., & Hicks, K. C. (2020a). Addressing the Advocacy Coalition Framework's neglect of policy design. Policy Sciences, 53(4), 691-708. https://doi.org/10.1007/s11077-020-09401-3

Pierce, J., DiSalvo, C., & Sengers, P. (2020b). Framing design inquiry into the infrastructures of everyday life. Design Issues, 36(1), 4–17. https://doi.org/10.1162/desi_a_00573

Pisano, G. P. (2020). The risk of de-risking innovation: Optimal R&D strategies in pharmaceutical firms. Administrative Science Quarterly, 65(3), 561–593. https://doi.org/10.1177/0001839219856711

Radnor, Z. J., Holweg, M., & Waring, J. (2012). Public sector service operations management: A research agenda. International Journal of Operations & Production Management, 32(11), 1249-1272. https://doi.org/10.1108/01443571211274595

Rieman, J. (1993). The diary study: A workplace-oriented research tool. Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (pp. 331-338). Association for Computing Machinery. https://doi.org/10.1145/169059.169255

Romme, A. G. L., & Meijer, A. (2020). Applying design science in public policy and administration research. Policy & Politics, 48(1), 149-165. https://doi.org/10.1332/030557319X15613699981234

Rowe, A. (2019). Rapid impact evaluation. Evaluation, 25(4), 496-513. https://doi.org/10.1177/1356389019870213

Sanderson, I. (2002). Evaluation, policy learning and evidence‐based policy making. Public Administration, 80(1), 1–22. https://doi.org/10.1111/1467-9299.00292

Schwartz, P. (1996). The art of the long view: Planning for the future in an uncertain world. Doubleday Business.

Shafir, E. (Ed.). (2013). The behavioral foundations of public policy. Princeton University Press.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. Education for Information, 22(2), 63-75. https://doi.org/10.3233/EFI-2004-22201

Shiffman, J. (2008). Generating political priority for newborn survival in five countries. Health Policy and Planning, 23(6), 339–356. https://doi.org/10.1093/heapol/czn024

Stickdorn, M., Hormess, M. E., Lawrence, A., & Schneider, J. (2018). This is Service Design doing: Applying Service Design thinking in the real world. O'Reilly Media.

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness. Yale University Press.

Timeus, K., & Gascó, M. (2018a). Increasing innovation capacity in city governments: Do innovation labs make a difference? Journal of Urban Affairs, 40(7), 992–1008. https://doi.org/10.1080/07352166.2018.1431078

Timeus, K., & Gascó, M. (2018b). Increasing the effectiveness of co-creation in public innovation labs. Public Management Review, 20(11), 1654–1672. https://doi.org/10.1080/14719037.2018.1424369

Tõnurist, P., & Hanson, A. (2020). Anticipatory innovation governance: Shaping the future through proactive policy making (OECD Working Papers on Public Governance, No. 44). OECD Publishing. https://doi.org/10.1787/cce14d80-en

Tõnurist, P., Kattel, R., & Lember, V. (2017). Innovation labs in the public sector: What they are and what they do? Public Management Review, 19(10), 1455–1479. https://doi.org/10.1080/14719037.2017.1287939

UK Policy Lab. (2019, July 3). Using prototypes in policy making. Open Policy Making Blog. https://openpolicy.blog.gov.uk/2019/07/03/using-prototypes-in-policy-making/

UK Policy Lab. (2021, January 30). Sprint before you walk: Fail faster. Public Policy Design Blog. https://publicpolicydesign.blog.gov.uk/2021/01/30/sprint-before-you-walk-fail-faster/

van Buuren, A., Lewis, J. M., & Puttick, B. (2020a). Understanding design practices in policy contexts: A systematic review of the literature. Policy Sciences, 53(4), 625–649. https://doi.org/10.1007/s11077-020-09398-5

van Buuren, A., Voorberg, W., & Bekkers, V. (2020b). Why do public organizations engage in co-design? Understanding the role of organizational characteristics and contextual factors. Public Money & Management, 40(1), 39-48. https://doi.org/10.1080/09540962.2019.1696745

Vedung, E. (1997). Public policy and program evaluation. Transaction Publishers.

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. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), New approaches to evaluating community initiatives: Concepts, methods, and contexts (pp. 65–92). Aspen Institute.

Wells, P. (2023). The uncertain future of government innovation labs. Apolitical. https://apolitical.co/solution-article/en/the-uncertain-future-of-government-innovation-labs

Wilkinson, A., & Kupers, R. (2013). Living in the futures: How scenario planning can help decision-makers to explore uncertainty and embrace complexity. Harvard Business Review, 91(5), 118–127.

Williams, B., & Hummelbrunner, R. (2010). Systems concepts in action: A practitioner's toolkit. Stanford University Press. https://doi.org/10.1515/9780804776554

APPENDICES

Appendix 1. Survey Questions

(As presented - in Latvian)

Tiesību un politikas jaunrades sprintu rezultātu izvērtēšana

Mērķauditorija -LatvijasRepublikas valsts pārvaldes un pašvaldību vajadzībām organizētu inovācijas un dizaiana sprintu dalībnieki.

Paredzamais aizpildīšanas laiks - līdz 5minūtēm

Aptaujasmērķis -iegūt informāciju par inovācijas un dizaina sprintu rezultātu izvērtēšanu

Anketētājs -Jānis Ķesa, Pakalpojumu dizaina stratēģijas un inovācijas (Erasmus Mundus - Latvijas mākslas akadēmija, Igaunijas mākslas akadēmija, Lapzemes universitāte) programmas maģistrants

Informācijas apstrāde un izmantošana - iegūtā informācija tiks lietota tikai akadēmiskās izpētesmērķiem

Sensitīva informācija - netiek pieprasīta

*Indicatesrequiredquestion

1. 1.Vai esat piedalījies/piedalījusies inovācijas vai dizaina sprintā? *

Mark only one oval.

- Jā
- Nē

Sprinta būtība

2. 2.Sprinta, kurā piedalījāties,rezultāts paredzēja: *

Mark only one oval.

- izmaiņas tiesību aktos, jaunu tiesību aktu vai vispārējo administratīvo aktu radīšanu
- izmaiņas vai jaunradi stratēģiskās vai politikas plānošanas dokumentos
- citu plašai sabiedrības daļai saistošu dokumentu vai praktisku risinājumu, kam ir faktiski saistošs raksturs, radīšanu vai grozīšanu
- nevienu no augstāk minētajiem risinājumiem

Sprinta rezultāts un tā izvērtēšana

Turpmāk ar vārdu "sprints" tiek apzīmēts tiesību vai politikas jaunrades sprints (dizaina vai inovācijas sprints), kurā esat piedalījies/piedalījusies.

3. 3.Vai sprinta rezultāts tika ieviests praksē? *

Mark only one oval.

- Jā, tika ieviests tāds, kā paredzēja sprinta iznākums
- Jā, tika ieviests, taču tikai daļēji vai ar korekcijām
- Nē, rezultāts neieguva lēmuma pieņēmēju atbalstu
- Other: _____

4. 4.Vai sprinta, kurā piedalījāties, rezultāts tika pārbaudīts ar lietotāju? *

Mark only one oval.

- Jā, ar pietiekoši nozīmīgu lietotāju skaitu, lai pārbaudes rezultātu uzskatītu par objektīvu sabiedrības viedokļa atspoguļojumu
- Jā, ar atsevišķiem lietotājiem, kas rada aptuvenu iespaidu par iespējamo sabiedrības viedokli kopumā
- Jā, lietotāji ir pati sprinta komanda - tās pārstāvētā institūcija
- Nē, rezultāta pārbaude ar lietotāju nebija iespējama dēļ risinājuma būtības
- Nē, rezultāta pārbaude ar lietotāju nebija nepieciešama dēļ risinājuma būtības
- Nē, rezultāta pārbaude ar lietotāju netika veikta, laika vai citu resursu trūkuma dēļ

5. 5.Vai sprinta,kurā piedalījāties, rezultātu izvērtējāt sprinta komandā pirms sprinta beigām? *

Mark only one oval.

- Jā, pārrunājām gala risinājumu vispārīgas diskusijas formātā
- Jā, pārspriedām, taču tikai kontekstā ar definēto problēmjautājumu
- Jā, tika izmantota metodoloģija vai rīki rezultāta izvērtēšanai
- Nē, uzticoties sprinta metodei, risinājums netika kā īpaši vērtēts vai pārspriests
- Nē, rezultāta būtība neprasīja risinājuma izvērtēšanu
- Nē, rezultātu izvērtēja lietotājs
- Nē, rezultātu izvērtēja "lēmuma pieņēmējs"

6. 6.Vai uzskatāt, ka sprinta komandai pašai būtu jāizvērtē paveiktais pirms sekojošajiem ieviešanas procesiem? *

Mark only one oval.

- Jā
- Nē, to jādara lietotājam
- Nē, to jādara kādam, kas nav bijis iesaistīts sprintā
- Nē, sprinta rezultāts ir tikai pirmais solis risinājuma virzienā
- Other: _____

7. 7.Kādi kritēriji būtu jāvērtē katra tiesību vai politikas jaunrades sprinta rezultāta izvērtēšanā? *

_____

8. 8.Vai uzskatāt, ka sprinta iznākuma izvērtēšanai sprinta komandas ietvaros būtu lietderīgi izmantot tam paredzētu ātras izvērtēšanas metodoloģiju - rīku? *

Mark only one oval.

- Jā
- Nē

9. 9.Pārstāvētā institūcija *

_____

Appendix 2. Interview Questions (For Policy Design Sprint Leaders)

(As asked - in Latvian)

SEMI-STRUKTURĒTAS INTERVIJAS JAUTAJUMI DIZAINA SPRINTU KOMANDU LĪDERIEM:

- Iepazīšanās;
- Atgādinājums par konfidencialitāti un tiesībām kontekstā ar ierakstu un izpēti;

JAUTĀJUMI

1. Kā definēt “policy design” sprinta veiksmi jeb izdošanos?

2. Kādi kritēriji, jūsuprāt, ir vissvarīgākie, novērtējot politikas dizaina sprintu rezultātus?

3. Kādi ir “policy design” sprinta iznākuma izvērtēšanas galvenie izaicinājumi?

4. Kā strukturētāka novērtēšanas metodoloģija varētu uzlabot nākotnes sprintu efektivitāti un rezultāta praktisku ieviešanu?

Apendix 3. Prototype “Survey Form of “Better Outcome” tool”

Better Outcome - A PolicyDesign Sprint Outcome Evaluation Form (Prototype)

A test form to validate the tool proposed by SDSI master's student JānisKesa within the process of research dedicated to his master's thesis, "Evaluating Policy Design Sprint Outcomes:A Proposition for a Novel Tool".

Filling out this test form does not require disclosing any sensitive data orinformation on actual sprints; it is designed for a "walkthrough" experience and validation with experts in the field.

*Indicates required question

1. Q1. Does your sprint outcome address the "How Might We?" question defined during the sprint? *

Mark only one oval.

1 2 3 4 5

Not at all ◯ ◯ ◯ ◯ ◯ Yes,totaly

2. Q2. Does the "How Might We?" question address the initial problem that defined the need for this spint? *

Mark only one oval.

1 2 3 4 5

Not at all ◯ ◯ ◯ ◯ ◯ Yes,totaly

3. Q3. Would you agree that the solution proposed by your sprint outcome is feasible - it can be realistically implemented within existing legal, political, financial, organisational and other constraints? *

Markonlyoneoval.

1 2 3 4 5

Not at all ◯ ◯ ◯ ◯ ◯ Yes.totaly

4. Q4.Who should be responsible for taking care of next steps so the outcome is implemented? *

EXISTING SOLUTION (the *STATUS QUO*)

The next 8 questions will be dedicated to setting a benchmark—a reference point that will serve as a comparison for the new alternative, the solution that is replacing the existing one,the solution that is the outcome of your Policy Design Sprint.

5. 1A. Does the existing solution work well within the legal system and its current architecture?*

Mark only one oval.

1 2 3 4 5

It totaly does not ◯ ◯ ◯ ◯ ◯ It pefectly does

6. 2A.Does the existing solution ensure justice and balance between different stakeholders?*

Mark only one oval.

1 2 3 4 5

It totaly does not ◯ ◯ ◯ ◯ ◯ It pefectly does

7. 3A.Does the existing solution respect the complexity of other systems apart from legal(social, natural, traditional values etc.)? *

Mark only one oval.

1 2 3 4 5

It totaly does not ◯ ◯ ◯ ◯ ◯ It pefectly does

8. 4A.Is the existing solution ready to face the future, and can it be considered future-ready?*

Mark only one oval.

1 2 3 4 5

It is totaly not ◯ ◯ ◯ ◯ ◯ It pefectly is

9. 5A. Does the existing solution address a genuine need and provide value to the public or specific target groups? *

Mark only one oval.

1 2 3 4 5

It totaly does not ◯ ◯ ◯ ◯ ◯ It pefectly does

10. 6A.Is the existing solution clear and understandable for the part of society it is relevant to?*

Mark only one oval.

1 2 3 4 5

It is totaly not ◯ ◯ ◯ ◯ ◯ It pefectly is

11. 7A.Does the existing solution reflect an understanding of end-user behavior and promote desired actions or outcomes? *

Mark only one oval.

1 2 3 4 5

It totaly does not ◯ ◯ ◯ ◯ ◯ It pefectly does

12. 8A.Does the existing solution create an administrative burden for the state that is viewed as "light"? *

Markonlyoneoval.

1 2 3 4 5

It's very heavy ◯ ◯ ◯ ◯ ◯ It is very "light"

PROPOSED SOLUTION (the SPRINT OUTCOME)

The next 8 questions will be dedicated to testing the outcome of your sprint the solution that is proposed to replace the existing one will be evaluated against the same criteria as the *StatusQuo* (previous section).

13. 1B.Would the proposed solution work well within the legal system and its current architecture? *

Mark only one oval.

1 2 3 4 5

It totaly would not ◯ ◯ ◯ ◯ ◯ It pefectly would

14. 2B.Would the proposed solution ensure justice and balance between different stakeholders? *

Mark only one oval.

1 2 3 4 5

It totaly would not ◯ ◯ ◯ ◯ ◯ It pefectly would

15. 3B.Does the proposed solution respect the complexity of other systems apart from legal (social, natural, traditional values etc.)? *

Mark only one oval.

1 2 3 4 5

It totaly does not ◯ ◯ ◯ ◯ ◯ It pefectly does

16. 4B.Would the proposed solution be ready to face the future,and can it be considered future-ready? *

Mark only one oval.

1 2 3 4 5

It totaly would not ◯ ◯ ◯ ◯ ◯ It pefectly would

17. 5B.Does the proposed solution address a genuine need and provide value to the public or specific target groups? *

Mark only one oval.

1 2 3 4 5

It totaly does not ◯ ◯ ◯ ◯ ◯ It pefectly does

18. 6B.Would the proposed solution be clear and understandable for the part of society it is relevant to?

Mark only one oval.

1 2 3 4 5

It totaly would not ◯ ◯ ◯ ◯ ◯ It pefectly would

19. 7B.Does the proposed solution reflect an understanding of end-user behaviour and promote desired actions or outcomes? *

Mark only one oval.

1 2 3 4 5

It totaly does not ◯ ◯ ◯ ◯ ◯ It pefectly does

20. 8B.Would the proposed solution create an administrative burden for the state that can be viewed as "light"? *

Mark only one oval.

1 2 3 4 5

It will be very heavy ◯ ◯ ◯ ◯ ◯ It will be very "light"

Appendix 4. Consent to participate in a research interview

sd—si

LAPIN YLIOPISTO
UNIVERSITY OF LAPLAND

Consent to take part in

Voice (and/or Video) recording of an interview for research purposes

This study is a part of SDSI Master Project studio which has been approved by the University of Lapland and SDSI (www.sdsi.ma).	Check if you agree with the next statement
I confirm that I have read, or had it explained to me on the date of __/04/2025 explaining the above research project, and I have had the opportunity to ask questions about the project.	
I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences. In addition, should I not wish to answer any particular question or questions, I am free to decline.	
I agree to the interview being audio- and video-recorded during the interview.	
I understand that members of the research team may have access to my anonymised responses in the video. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research. I understand that recorded voice & videos are being used as one of the data collection methods. Full anonymity cannot be guaranteed on behalf of the other interview participants.	
I understand that the data collected from me may be stored under the Data Protection Act (1050/2018) and used in relevant future research in an anonymised form.	
I agree to take part in the above research project and will inform the researchers should my contact details change.	

Name of participant		Signature
Email address (optional)		Date

Contact person at the University of Lapland (Name / Position) & E-mail address

Jānis Ķesa (Researcher)	jkesa@ulapland.fi
Kiwoong Nam (Supervisor)	kiwoong.nam@ulapland.fi