Policy Design and Practice

ISSN: 2574-1292 (Online) Journal homepage: www.tandfonline.com/journals/rpdp20

# Developing a design-led evaluation model for citizen-centered public services: a South Korean case

Ahmee Kim & Hyung Jun Park

**To cite this article:** Ahmee Kim & Hyung Jun Park (02 Jul 2025): Developing a design-led evaluation model for citizen-centered public services: a South Korean case, Policy Design and Practice, DOI: 10.1080/25741292.2025.2523104

**To link to this article:** https://doi.org/10.1080/25741292.2025.2523104



Published online: 02 Jul 2025.

POLICY DESIGN AND PRACTICE
https://doi.org/10.1080/25741292.2025.2523104

Routledge
Taylor & Francis Group

RESEARCH ARTICLE



# Developing a design-led evaluation model for citizen-centered public services: a South Korean case

Ahmee Kim[a] and Hyung Jun Park[b]

[a]Institute of Governance & Policy Evaluation, Sungkyunkwan University, Seoul, South Korea;
[b]Department of Public Administration & Graduate School of Governance, Sungkyunkwan University, Seoul, South Korea

**ABSTRACT**

Public services have long been evaluated based on service quality and citizen satisfaction. However, as public services are increasingly understood as being co-created within broader service ecosystems, and as "co-" practices gain prominence in the public sector, a new approach to evaluation is needed. This study proposes a preliminary design-led model for citizen-centered public service evaluation, emphasizing stakeholder participation and systemic perspectives, drawing on a Korean case. The model's value is validated through expert opinions from public administration and design fields. Findings highlight the model's potential to address complex service ecosystems and promote long-term change, while identifying challenges such as context-specific application and stakeholder capabilities.

**ARTICLE HISTORY**
Received 18 March 2025
Accepted 12 June 2025

**KEYWORDS**
Public service evaluation; design-led evaluation; evaluation model; public service ecosystem; design for policy

## 1. Intro

Governments worldwide have shifted from viewing citizens as passive recipients to active co-producers of public services (Baredes 2022). In South Korea, the Ministry of the Interior and Safety launched the 2021 Citizen Participation Activation Promotion Plan to institutionalize citizen participation by strengthening administrative capacity and expanding engagement throughout policymaking (Ministry of the Interior and Safety 2021).

Citizen-centred approaches are considered more transparent and effective than those based on policymakers' assumptions, improving satisfaction and trust in government (The World Bank 2018). Digital technologies further enable swift responses to citizen needs and real-time feedback collection (OECD 2019).

This study proposes a design-led model for citizen-centered public service evaluation, addressing the need for evaluation frameworks that reflect co-creation within service ecosystems. Design-led evaluation employs design methodologies (e.g. journey

**CONTACT** Hyung Jun Park hjpark72@skku.edu Department of Public Administration & Graduate School of Governance, Sungkyunkwan University, Seoul, South Korea.

mapping, prototyping) to facilitate stakeholder collaboration and qualitative insights, while citizen-centered evaluation prioritizes citizens' experiences and co-creation throughout the service process. Design methodologies operationalize citizen-centeredness by enabling iterative, participatory, and user-focused evaluation processes. This study's novelty lies in integrating these approaches within South Korea's unique governance context, characterized by centralized, top-down evaluation systems (Yang and Torneo 2016). Unlike design-led evaluations in sectors like healthcare, which focus on specific service contexts (e.g. patient experiences), this model adopts a broader service ecosystem perspective, applicable to diverse public services.

Drawing on literature on public sector design and service evaluation, this study develops a preliminary model and examines its value through expert opinions. Two research questions guide the study: (1) What value do experts perceive in the design-led evaluation model? (2) What challenges exist for its further development and implementation? The paper reviews traditional evaluation approaches, explores opportunities of design-led evaluation, details the research design, and presents findings, including an applied example of the model.

## 2. Approaches to evaluating public services

Public policy evaluation began in the 1960s in the U.S. and was later adopted in Sweden, Canada, and Germany (Derlien 1999). In the 1980s and 1990s, reforms emphasized citizen-oriented services (Van de Walle 2018). By the late 1990s, public service markets were liberalized, shifting from state-led to citizen demand-driven models (Van de Walle 2018).

During this period, service quality and citizen satisfaction became dominant evaluation criteria, influenced by private-sector quality management (Brysland and Curry 2001). As outsourcing expanded, the idea of delivering "best value" through competition gained traction. Scholars also linked citizen satisfaction to trust in government, showing that low satisfaction leads to negative evaluations and reduced support for public programmes (Van De Walle and Bouckaert 2003; Welch, Hinnant, and Moon 2004).

The study of service quality originated in the private sector, where it was considered harder to evaluate objectively than product quality (Parasuraman, Zeithaml, and Berry 1988). It was defined as the gap between customer expectations and performance (Grönroos 1984; Parasuraman, Zeithaml, and Berry 1985), requiring ongoing monitoring and management (Carman 1990).

Meanwhile, citizen satisfaction is a complex concept that encompasses expectations, experiences, and prior attitudes (Van de Walle 2018). Improvements in service quality do not necessarily translate into higher citizen satisfaction. With the rise of customer orientation, surveys assessing customer satisfaction have been conducted in many countries since the 1990s (Van de Walle 2018). In South Korea, such surveys were first introduced in 1999 and have been regularly conducted ever since (Ministry of Economy and Finance 2012).

While public service evaluations based on service quality and citizen satisfaction remain relevant today, research on service has shifted its focus toward service experience (Sundbo 2015). Service experience is conceptualized in terms of overall value

and higher-order objectives. For instance, in the context of travel services, customers do not merely seek accommodations or flight reservations but also value-creating experiences, such as enjoyable moments with family (Klaus and Maklan 2013). More importantly, the value creation is viewed as co-created within the service ecosystem through collaboration between providers and customers (Vargo, Akaka, and Vaughan 2017).[1]

This shift in perspective has also influenced public services. Osborne (2018) argues that co-creation in public services occurs inherently, even if citizens are unaware of it. Furthermore, the global proliferation of public sector innovation (PSI) labs exemplifies increasing adoption of "co-" modes of collaboration with stakeholders in the public sector (McGann, Blomkamp, and Lewis 2018).

Despite the growing emphasis on a service ecosystem perspective in public service research, studies evaluating public services from this lens remain relatively scarce, particularly compared to research on co-practices in the early stages of the policy process (i.e. planning and implementation). Both the practice and study of co-evaluation in public interventions remain largely unexplored (Mckenna 2021).

## 3. Potentials of design-led evaluation

With the rise of PSI labs, design has been recognized as a method for public sector innovation (Cole 2022; McGann, Blomkamp, and Lewis 2018). This trend also opens up possibilities for public service evaluation to generate meaningful changes through the use of design approaches.

Firstly, design approaches such as user-centred and co-design help deliver services more responsive to user needs and reduce distrust in government (Blomkamp, 2018; Mintrom and Luetjens, 2016). Depending on the nature of the problem and the organizational context, various forms of citizen participation are possible – design *for*, *with*, or *by* citizens (Junginger 2017). These participatory approaches can be applied either partially or fully across different stages of the policy process (Nabatchi, Sancino, and Sicilia 2017). Embedding such participatory practices can transform public service evaluation.

Secondly, design approaches – particularly those rooted in human-centred design – do not take a given problem at face value. Instead, they seek to redefine the problem by identifying its underlying causes through iterative cycles of divergence and convergence. They incorporate user perspectives throughout and employs techniques such as prototyping and reflective practice to enable ongoing validation and feedback (Norman 2013). In the policy context, Mintrom and Luetjens (2016) describe this as environmental scanning, arguing that design enables policy actors to revisit and critically examine long-standing assumptions about public problems. Accordingly, design can support the evaluation of whether public issues have been appropriately understood and whether solution-seeking processes have been adequately explored.

Thirdly, as human-centred design expands into the realm of addressing complex societal issues, the notion of the "user" has also broadened – from the end-user of a service to a wider ecosystem of stakeholders related to the issue (van der Bijl-Brouwer 2016). In addition, design practices now often integrate interdisciplinary

methods such as ethnography, systems thinking, and sociology. Alongside these developments, PSI labs – which lead public sector design initiatives – are increasingly engaging in co-design with community partners to address pressing social challenges (Cole and Hagen 2024). While some scholars have criticized PSI labs for focusing on short-term outcomes (Olejniczak et al. 2020), Cole and Hagen (2024) propose design integrated with the theory of change[2] as a structured approach to enable systemic and long-term transformations. This suggests that when integrated with interdisciplinary approaches, design holds potential as a valuable means to evaluate both short-term and long-term impacts in public services, particularly by understanding user and stakeholder experiences.

Fourthly, design has been embraced in the public sector as a mode of experimentalism (Kimbell and Bailey 2017). Based on their analysis of 20 PSI labs in 16 countries, Olejniczak et al. (2020) argued that experimentation through prototyping in the policy process can bridge the gap between policies and user needs by creating small iterative cycles of designing, testing, and adapting within the larger policy cycle. Furthermore, as design is increasingly used for long-term transformation, prototyping is understood not only as a tool for short-term learning, but also as a form of agonistic space – a setting in which stakeholders surface dilemmas and explore new possibilities through extended, iterative experimentation (Hillgren, Seravalli, and Emilson 2011). Addressing complex public challenges often requires long-term processes where innovation emerges through the co-evolution of stakeholder relationships, knowledge, and practices (van der Bijl-Brouwer, Kligyte, and Key 2021). In this sense, evaluating how learning and feedback unfold within design processes is critical to understanding whether, and how, innovation arises in response to complexity.

Taken together, these perspectives suggest that design approaches hold considerable potential for advancing public service evaluation – particularly in relation to stakeholder participation, problem exploration, user and stakeholder experiences, learning and feedback processes, and both short- and long-term impacts. These considerations informed the development of the evaluation dimensions and items proposed in this study, which will be elaborated in the following section.

## 4. Method

This study aimed to develop a preliminary model for evaluating public services using design approaches. Drawing on existing literature on design in the public sector and service evaluation, we constructed the model and examined its value through expert input using the Delphi method. The research questions were: (1) What value do experts perceive in the design-led evaluation model? (2) What challenges exist for its further development and implementation?

### *4.1. Research context: Public service evaluations in South Korea*

In South Korea, public services are evaluated under the Framework Act on Government Performance Evaluation, which covers central and local governments, affiliated organizations, and public institutions. These evaluations are conducted

annually, focusing on specific topics or overall performance (Office for Government Policy Coordination n.d)

While influenced by U.S. models, Korea's evaluation system retains a top-down structure shaped by Confucian traditions, centralized governance, and a relatively short democratic history (Yang and Torneo 2016; Jung 2024). Jang and Lee (2024) identified a perception gap between evaluating agencies and those being evaluated. The former view evaluations as tools for performance verification and improvement, while the latter perceive them mainly as mechanisms of control and inter-agency comparison. Because evaluation results are often reduced to rankings and numeric scores, they have limited effectiveness in identifying problems or generating actionable insights. These studies underscore the challenges of producing meaningful feedback for local and service-level public ecosystems under the current evaluation system.

## 4.2. Design of a preliminary evaluation model

### 4.2.1. Foundational concepts

The model is grounded in three foundational concepts, which underpin its structure and differentiate it from traditional evaluation approaches (see Table 1 below). The first concept is the service ecosystem. Public services are understood as networks where multiple stakeholders (e.g. citizens, public officials, NGOs) co-create value through dynamic interactions (Vargo, Akaka, and Vaughan 2017). This informs the stakeholder participation and organizational/ecosystem capabilities dimensions. The second concept is systems thinking, a holistic approach to understanding complex interconnections in societal issues (Morçöl 2005). This shapes the problem definition and end user experience dimensions. The third concept is the theory of change which is a structured roadmap linking short-, medium-, and long-term goals to achieve systemic impact (Cole 2022). This underpins the outcome/impact and learning/feedback dimensions. These concepts ensure the model captures the complexity of public services, promotes stakeholder collaboration, and supports long-term change.

**Table 1.** Comparison of traditional and design-led evaluation criteria.

| Criteria | Traditional Evaluation (e.g. SERVQUAL[a], American Customer Satisfaction Index [ACSI][b]) | Design-Led Evaluation (Proposed Model) |
|---|---|---|
| Focus | service quality, citizen satisfaction | co-created value, ecosystem perspective |
| Methodology | quantitative surveys, standardized metrics | qualitative, participatory design tools (e.g. journey mapping) |
| Stakeholder Role | passive respondents | active co-producers across service stages |
| Time Orientation | ex-post, short-term | real-time, long-term systemic change |
| Key Dimensions | reliability, responsiveness, assurance, empathy, tangibles (SERVQUAL); expectations, quality, value (ACSI) | stakeholder participation, problem definition, organizational/ ecosystem capabilities, end user experience, outcomes/impact, learning/feedback |

[a]Parasuraman et al. (1990).
[b]Fornell et al. (1996).

### 4.2.2. Model structure

The model comprises six dimensions: stakeholder participation, problem definition, organizational and ecosystem capabilities, end user experience, outcomes and impact, and learning and feedback (Figure 1). Each dimension includes items (Table 2), designed to be selectively applied based on service context, following the PDCA (Plan-Do-Check-Act) cycle for continuous improvement.

To illustrate how the proposed evaluation model can be applied in practice, we present two hypothetical scenarios (also see Table 3). One concerns a passport service, and the other concerns dementia care.In both cases, the evaluator collaborates with relevant stakeholders and follows a process similar to developing a public service, as the evaluation team must first understand the problem in order to assess whether the service adequately addresses it. As shown in the table below, we assume that depending on the complexity of the issue, the evaluation team will apply different sets of items from the model and select appropriate design tools to support the assessment.

In South Korea, the passport service is a relatively simple and short-term problem, where applications are submitted online or offline and passports are collected in person within approximately two weeks. In such cases, the evaluation team can map the user journey - a description of how users interact with the service across different touchpoints (Service Design Tools n.d) - and assess the short-term outcomes, focusing on how smoothly the service is delivered. In contrast, dementia care services vary across regions in Korea in terms of medical and care infrastructure and the stakeholders involved. Dementia care requires long-term and systemic responses. Therefore, the evaluation team should identify the long-term vision of the local community and assess how efforts are being made in the short, medium, and long term to realize that vision. In this context, we suggest using design tools integrated with systems thinking.

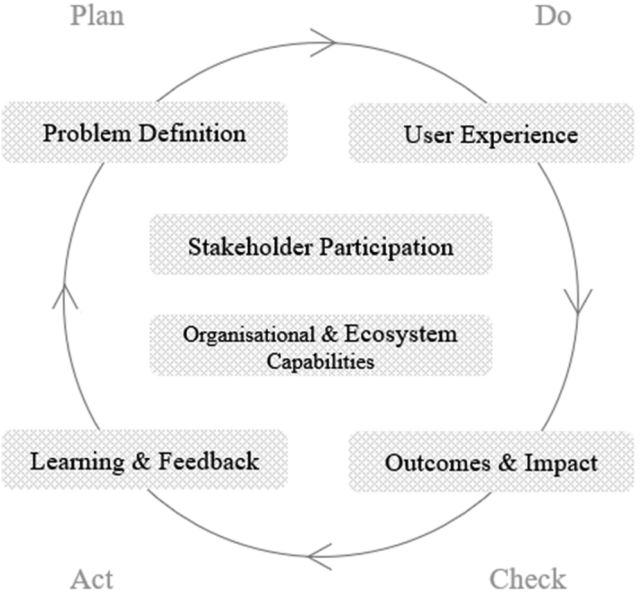


**Figure 1.** A preliminary design-led evaluation model.

**Table 2.** Items in six dimensions of the model[a].

| Dimension | | Items |
|---|---|---|
| **Stakeholder Participation** | | a) Who were the stakeholders? (e.g. identities, issue relevance, power or knowledge gaps, and resources used)<br>b) How and how often did they participate? (e.g. surveys, interviews, workshops)<br>c) Which service stages did they engage in? (planning, implementation, evaluation, or action)<br>d) How effectively was participation facilitated? |
| **Problem Definition** | | a) What innovation methods were used to understand the problem?<br>b) Were solutions iteratively tested and transparently improved (e.g. publicly documented)?<br>c) Were short-, medium-, and long-term goals set based on insights (e.g. using a theory-of-change roadmap)?<br>d) Was the problem's spatial and temporal impact considered? |
| **Organizational & Ecosystem Capabilities** | | a) Do practitioners and stakeholders have the capability to understand and apply innovative methods?<br>b) Are opportunities (e.g. training, workshops, mentoring) provided to build their capabilities? |
| **End User Experience**[b] | Usability | a) Accessibility: Was it easy to access the service and information? Were digital and non-digital touchpoints well integrated?<br>b) Responsiveness: Was the service timely and tailored?<br>c) Innovativeness: Did the service include new ideas that benefited users? |
| | Affect | a) Aesthetic Quality: Were touchpoints simple, free of clutter, and visually satisfying?<br>b) Comfort: Were physical touchpoints pleasant and comfortable?<br>c) Reliability: Were staff professional, services consistent, and the experience trustworthy?<br>d) Linguistic Appropriateness: Did language at touchpoints make users feel welcomed and respected? |
| | User Value | a) Fairness: Were equal standards applied to all, and were support and benefits fairly distributed?<br>b) Integrity: Did staff follow ethical standards, and was decision-making transparent?<br>c) Inclusiveness: Were users' diverse social and physical backgrounds (e.g. language, culture, disabilities) considered?<br>d) Digital Security: Was personal data and cybersecurity properly managed? |
| **Outcomes & Impact** | | a) Did the service address users' short-term problems?<br>b) How did current outcomes contribute to long-term goals?<br>c) Did the service help address other public issues? If so, how? |
| **Learning & Feedback** | | a) How often was feedback from users and stakeholders collected?<br>b) Was the service modified based on feedback? (Single-loop learning)<br>c) Were core assumptions or goals reviewed and revised? (Double-loop learning)<br>d) Was the feedback process – collection, learning, and use – transparent? |

[a]This table presents the revised model reflecting expert feedback.
[b]The dimension *User & Stakeholder Experience* was renamed *End User Experience* based on expert input. Sub-items were informed by both expert insights and literature on usability, affect, and user value (Lee, Lee, and Choi 2018).

### *4.3. Validation through expert opinion collection informed by the Delphi method*

The Delphi method seeks to reach a consensus on policy, practice, or organizational decisions through structured communication among experts, allowing them to freely assess and comment under conditions of anonymity. This process typically involves

**Table 3.** Application of the model to two hypothetical cases with varying complexity[a].

| Case | Items evaluated | Tools used |
|---|---|---|
| Passport service | The focus of evaluation will be around the dimension of end user experience, single loop learning, and short-term outcome. | User journey mapping to find out how the user experience passport issuing process |
| Dementia care service | A wide range of dimensions will be considered, including problem definition, stakeholder participation, and long-term outcomes. | Design tools integrated with systems thinking (e.g. Systemic Design Toolkit by Design Council 2021) |

[a]These cases are illustrative only; item and tool selection may vary by evaluation context.

**Table 4.** Participant information.

| Participant | Field of expertise | Profession |
|---|---|---|
| P01 | Design | Practitioner (Design Agency) |
| P02 | Public Administration | Academic |
| P03 | Public Administration | Academic |
| P04 | Public Administration | Academic |
| P05 | Design | Academic |
| P06 | Public Administration | Academic |
| P07 | Design | Academic |
| P08 | Public Administration | Academic |
| P09 | Design | Practitioner (Design Agency) |
| P10 | Public Administration | Practitioner (Research Institute) |

both quantitative and qualitative approaches over multiple rounds to build consensus (Beiderbeck et al. 2021).

The Delphi questionnaires (see Appendix A) were distributed in two rounds, with responses collected in writing and supplemented by follow-up emails as needed. In the first round, experts were asked to critically assess the model's structure and items based on their professional or academic experience. Agreement was considered reached when more than 70% of participants (i.e. at least 7 out of 10) concurred on an item. In the second round, experts reviewed the revised model and responded to additional questions regarding its further development. Unlike the first round, this phase did not seek consensus but instead collected open-ended feedback. These qualitative insights were thematically analyzed and used to refine the final model and interpret expert perspectives. While the first round identified areas of agreement, the second round offered deeper understanding of points of contention and suggestions for improvement. This iterative process positions the study as a modified Delphi method rather than a standard one.

Experts were recruited from our academic network and included professionals and researchers specializing in public sector design or public administration. Selection was based on subject expertise and relevance to South Korea's public sector, with a balance of practitioners (3) and academics (7). While it was difficult to identify Korean experts with deep knowledge across design, public administration, and evaluation, we prioritized a balance between practitioners and researchers. Most public administration experts were affiliated with academia but were also actively involved in government performance evaluation. Responses were collected in writing, with follow-up emails for clarification. Citizens were not included, as the focus was on theoretical validation; broader stakeholder engagement is planned in future research. Table 4 summarizes their profiles.

The data analysis was guided by the two research questions. Initially, the data were segmented and coded with an open-coding approach in ATLAS.ti software. Codes were then grouped iteratively to form themes that addressed the research questions (see Appendix B for an example). The first author led the analysis, which was then validated through independent reviews by the second author to ensure reliability. A full list of codes and themes is available in Appendix C.

## 5. Findings

Experts described the new evaluation model as contrasting to conventional evaluation approaches and embracing complexity of public services. As to the challenges in its development/implementation, several challenges were identified.

### 5.1. Contrast to conventional evaluation approaches

Experts perceived the new evaluation model as contrasting with conventional evaluation approaches. The current public service evaluation was described as focused on short-term problem-solving, formalistic, of bureaucratic perspective, and ex-post.

The focus on short-term problem-solving meant that the current public service system prioritizes the achievement of short-term goals due to structural constraints such as annual fiscal year cycles of public organizations and terms of political leadership (P3, P4). Being formalistic indicated that the original purpose of evaluation – providing feedback to improve services – was not being fulfilled. Instead, it was reduced to procedural tasks, such as counting how many times stakeholder meetings has been held (P4, P8, P9, P10). Bureaucratic perspective was described as "detached from reality," implying that the evaluations failed to reflect the actual needs of stakeholders (P4). Ex-post referred to evaluations being executed according to annual government performance evaluation cycle (P4).

In contrast, the new evaluation model was perceived as supporting long-term problem-solving (P3), being based on qualitative data (P8, P9), incorporating stakeholder needs or enabling direct participation, and allowing real-time feedback (or as needing to do so) (P2, P3, P4). While the rest are self-explanatory, real-time feedback was described as not being bound by the government performance evaluation cycle but instead enabling stakeholders to provide feedback at any time and stage of the public service process (P3, P4, P6).

### 5.2. Embracing the complexity of public services

Several experts perceived public services as inherently complex processes. They described public services as "the intricate and multifaceted outcomes of policy intent, budget formulation, task execution, and operation/maintenance" (P5) and as the interactions of diverse and broad-ranging stakeholders (P1, P2). P4 also noted that societal issues are directly or indirectly interconnected. The new evaluation model was recognized (or as needing to be) as an approach that considers this complexity

(P2, P7, P10). Some expressed that the model should adopt a service-ecosystem perspective in its implementation (P7, P9).

### 5.3. Different applications depending on the context of the problem

From the first round of the expert opinion collection, a concern was raised about the uniform application of the evaluation model. P5 pointed out that presenting an "evaluation standard" risks of applying it uniformly across all public services. We had also recognized this concern during the model's designing process, as public service problems vary significantly in complexity – from fixing streetlights to addressing large-scale challenges like climate change.

Other experts also agreed and suggested that the evaluation model's items could be applied differently depending on the nature and type of the problem (P5, P6, P8, P10). P8 proposed developing a set of common items and specialized items, where the former could be applied to relatively simple service problems, while both the common and specialized items could be applied to complex service problems.

However, challenges were also identified in developing evaluation items tailored to varying complexities of public service problems. These challenges included the potential costs involved in the development process, the need to ensure that the model is easy for evaluators and stakeholders to use and understand, and the difficulty of comparing public service cases when tailored evaluations use distinct items (P4).

### 5.4. Tension between autonomy and standardization

Experts had diverse opinions regarding the segmentation and application of evaluation items, revealing a tension between autonomy and standardization in this discussion. P5 suggested that the evaluation model should not become a rigid standard but instead present diverse criteria, allowing stakeholders to select appropriate ones based on the context of their problems. Similarly, other experts viewed the model as serving as a "minimum benchmark" (P9) or a "guideline" (P10).

In contrast, some experts suggested further refining the evaluation items by developing more detailed indicators or formalizing them. For instance, P4 and P8 proposed creating indicators to assess stakeholder participation, such as the relevance of stakeholders to the problem, the mode of participation (in-person or virtual), the frequency of participation, and the extent of meaningful engagement. P8 recommended formalizing the model by introducing tailored indicators for specific groups. However, P10 expressed concerns about excessive segmentation and quantification of evaluation items.

### 5.5. Continuous improvement and adaptation of the evaluation model, and its obstacles

5 expressed concern that this evaluation model could become formalistic like conventional evaluations – a view shared by most experts (P2, P3, P4, P6, P7, P8, P9, P10). To address this, several experts (P2, P7, P9, P10) proposed ongoing monitoring

of the model. They emphasized the importance of establishing a feedback loop where evaluation results are used to improve public services and ensure that the loop remains active. P2 also highlighted the need to assess whether the model can adapt to changing environments and needs.

However, obstacles to such continuous improvement and adaptation were identified, namely job rotation[3] and the outsourcing of public services. P5 and P10 viewed job rotation as a barrier, as it weakens employee expertise and project continuity, leading to evaluations being seen as formalities. P5 also identified outsourcing as a challenge, arguing that it blurs accountability for learning and incorporating feedback. This lack of clear responsibility, according to P5, hampers the possibility of achieving double-loop learning.

### 5.6. Lack of capabilities in the public service ecosystem

For the proposed evaluation model to be effectively implemented, enhancing the capabilities of diverse stakeholders involved in public service evaluation was repeatedly mentioned.

First, the need to enhance evaluators' capabilities was emphasized, as qualitative evaluations may yield varying results depending on their expertise (P2, P9), raising concerns about the objectivity and reliability of the evaluation process. Second, organizational capability was highlighted, particularly the importance of fostering "creative confidence" – the openness to diverse, innovative ideas and continuous feedback – as well as the ability to apply service design principles and address complex problems (P2, P3, P9). Third, the importance of organizational leaders' understanding and support for new approaches to evaluation was noted (P3, P5). Lastly, given that public services are often delivered through outsourcing, P4 suggested that the capabilities of external stakeholders should also be enhanced.

## 6. Discussion and conclusions

This study proposed a preliminary design-led evaluation model for public services. Experts in the Korean public sector perceived the proposed design-led evaluation model as having the potential to accommodate the complexity of public services, promote long-term change, and enable stakeholder participation in contrast to conventional evaluation approaches.

However, challenges remain. Some reflect Korea's centralized evaluation culture, while others – such as a preference for quantification and lack of stakeholder capabilities – are more universal. Morçöl (2005) argued that complexity theory in public administration recognizes policy as a system that evolves with its environment, acknowledges its unpredictability, and highlights the need to embrace heuristics and qualitative approaches. From this view, the tension between autonomy and standardization reveals the challenges of adopting systems thinking in public service evaluation. Such resistance often stems from entrenched organizational cultures or limited awareness of its value (Apolitical 2024).[4]

This study also offers directions for future research. First, it serves as a starting point for empirical inquiry. Future studies could explore how the model applies to different public sector contexts. In doing so, involving a broader range of stakeholders – including public officials, academics, NGOs, and citizens – will be important. As the model aspires to support citizen-centered and "co-" evaluation, citizen participation in refining the model is crucial. Due to time constraints, such engagement was not possible in this study.

Second, this study was conducted in the Korean context, where the evaluation of public programmes tends to be authoritarian. The practice of external experts evaluating public programmes at specific intervals under the central government's direction is not unique to Korea. However, evaluation systems, local autonomy, and evaluation-related capabilities vary significantly between countries and regions. For instance, PSI labs of the Auckland City Government in New Zealand developed its own evaluation framework through long years of co-design practices with indigenous communities and local stakeholders. They argued that traditional government-led evaluation methods are limited in addressing complex social issues and supporting systemic change (Tarena et al. 2021). However, their evaluation approach is closely linked to the innovation capacity in the ecosystem. Future research is needed to explore how such a design-led model can be adapted to different institutional settings and cultural norms globally. For example, in a country like Korea, where public officials are subject to frequent job rotations, strengthening the capabilities of local stakeholder networks may be a more critical issue than in countries without such institutional arrangements.

This study's contribution lies not in the proposed evaluation model itself but in initiating discussions and setting the stage for future research. We explored the value of citizen-centered public service evaluation enabled by design approaches and discussed future research agendas and implementation challenges for its effective application. Insights from the Korean context may inform global discussions on integrating design into public service evaluation, especially in systems transitioning toward inclusive and ecosystem-based practices.

## Notes

1. Vargo and Lusch (2016) define service ecosystem as "a relatively self-contained, self-adjusting system of resource-integrating actors connected by shared institutional arrangements and mutual value creation through service exchange" (pp. 10-11). This highlights the need to consider multi-level factors shaped by social and cultural contexts when examining services.
2. The theory of change illustrates how an intervention or program achieves its goals by setting long-term outcomes and working backward to identify necessary preconditions through causal pathways. It provides a roadmap for monitoring and evaluation, while also supporting reflection, transparency, and consensus-building (Taplin et al. 2013). It is widely used to evaluate public interventions, including by the Government of Canada and the European Commission (Treasury Board of Canada Secretariat 2021; European Commission 2024).
3. Job rotation is a key aspect of the personnel system in Korean government organisations, intended to prevent stagnation and ensure an optimal service length for expertise and

efficiency. However, senior officials in Korea serve an average of one year, far shorter than the five-year OECD average, hindering expertise accumulation, disrupting continuity, and weakening accountability (Kim 2008).

4. In this study, systems thinking is not applied as a formal framework, but rather reflected in the recognition that public services are co-produced within complex service ecosystems, and that certain public service problems require long-term, multi-level perspectives to address effectively.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## ORCID

Ahmee Kim http://orcid.org/0000-0002-8678-0567

## References

Apolitical. 2024. "The Beginner's Guide to Systems Thinking." *Apolitical*. https://apolitical.co/solution-articles/en/the-beginners-guide-to-systems-thinking

Baredes, B. 2022. "Serving Citizens: Measuring the Performance of Services for a Better User Experience." OECD Working Papers on Public Governance Vol. 52.

Beiderbeck, D., N. Frevel, H. A. Von Der Gracht, S. L. Schmidt, and V. M. Schweitzer. 2021. "Preparing, Conducting, and Analyzing Delphi Surveys: Cross-Disciplinary Practices, New Directions, and Advancements." *MethodsX* 8: 101401. https://doi.org/10.1016/j.mex.2021.101401.

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.

Brysland, A., and A. Curry. 2001. "Perspectives Service Improvements in Public Services Using SERVQUAL." *Managing Service Quality* 11 (6): 389–401.

Carman, J. M. 1990. "Consumer Perceptions of Service Quality: An Assessment of the SERVQUAL Dimensions." *Journal of Retailing* 66 (1): 33–55.

Cole, L. 2022. "A Framework to Conceptualize Innovation Purpose in Public Sector Innovation Labs." *Policy Design and Practice* 5 (2): 164–182. https://doi.org/10.1080/25741292.2021.2007619.

Cole, L., and P. Hagen. 2024. "Scaling Deep through Transformative Learning in Public Sector Innovation Labs – Experiences from Vancouver and Auckland." *Public Management Review* 26 (7): 2094–2121. https://doi.org/10.1080/14719037.2023.2254776.

Derlien, H.U. 1999. "Genesis and Structure of Evaluation Efforts in Comparative Perspective." In *Program Evaluation and the Management of Government* (pp. 147–175). Routledge. https://doi.org/10.4324/9780429338380

Design Council. 2021. "The Systemic Design Toolkit." https://www.designcouncil.org.uk/our-resources/systemic-design-framework/

European Commission. 2024. "EU Support to Refugees in Türkiye: Theory of Change." https://neighbourhood-enlargement.ec.europa.eu/document/download/99c1bfa3-1853-42ff-ae3b-59622f7bdab1_en?utm_source=chatgpt.com

Fornell, C., M. D. Johnson, E. W. Anderson, J. Cha, and B. E. Bryant. 1996. "The American Customer Satisfaction Index: Nature, Purpose, and Findings." *Journal of Marketing* 60 (4): 7–18. https://doi.org/10.1177/002224299606000403.

Grönroos, Christian. 1984. "A Service Quality Model and Its Marketing Implications." *European Journal of Marketing* 18 (4): 36–44. https://doi.org/10.1108/EUM0000000004784.

Hillgren, P. A., A. Seravalli, and A. Emilson. 2011. "Prototyping and Infrastructuring in Design for Social Innovation." *CoDesign* 7 (3-4): 169–183. https://doi.org/10.1080/15710882.2011.630474.

Jang, H., and H. Lee. 2024. "A Study on the Institutionalization of the Cross-Agency Evaluation: Focusing on Evaluation Management, Utilization and Reporting." *Legislation and Policy* 16 (2): 153–182.

Jung, Y. 2024. "A Study on the Improvement of the Effectiveness of Feedback of Government Performance Evaluation." *The Journal of the Convergence on Culture Technology* 10 (3): 541–550.

Junginger, S. 2017. *Transforming Public Services by Design*. New York: Routledge.

Kim, K. H. 2008. "A Study on Job Rotation in Civil Service." *KDI Journal of Economic Policy* 30 (2): 61–97.

Kimbell, L., and J. Bailey. 2017. "Prototyping and the New Spirit of Policy-Making." *CoDesign* 13 (3): 214–226. https://doi.org/10.1080/15710882.2017.1355003.

Klaus, P., and S. Maklan. 2013. "Towards a Better Measure of Customer Experience." *International Journal of Market Research* 55 (2): 227–246. https://doi.org/10.2501/IJMR-2013-021.

Lee, H., K. H. Lee, and J. Choi. 2018. "A Structural Model for Unity of Experience: Connecting User Experience, Customer Experience, and Brand Experience." *Journal of Usability Studies* 11 (1): 8–34.

McGann, M., E. Blomkamp, and J. M. Lewis. 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.

Mckenna, D. 2021. "Co-Assessment through Citizens and Service Users in Audit, Inspection and Scrutiny." In *The Palgrave handbook of co-production of public services and outcomes*, edited by E. Loeffler & T. Bovaird, 451–467. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-53705-0_23

Ministry of Economy and Finance. 2012. "Customer Satisfaction Index of Public Institutions Declines for the First Time in 13 Years… Average Score 89.2. Korea Policy Briefing." https://www.korea.kr/briefing/pressReleaseView.do?newsId=155869404

Ministry of the Interior and Safety. 2021. "Participation as Part of Daily Life, Proposals as Policy: Expanding Public Participation in All Areas of Administration." Regulation Information Portal. https://www.better.go.kr/hz.blltn.BodoBlltnSl?brd_seq=16&bltn_seq=13355&link_mini_cd=1192000

Mintrom, M, and J. Luetjens. 2016. "Design Thinking in Policymaking Processes: Opportunities and Challenges." *Australian Journal of Public Administration* 75 (3): 391–402. 10.1111/1467-8500.12211.

Morçöl, G. 2005. "A New Systems Thinking: Implications of the Sciences of Complexity for Public Policy and Administration." *Public Administration Quarterly* 29 (3): 297–320. https://doi.org/10.1177/073491490502900303.

Mucz, D., and C. Gareau-Brennan. 2019. "Evaluating Customer Experience through Journey Mapping." *Partnership: The Canadian Journal of Library and Information Practice* 14 (1). http://doi.org/10.21083/partnership.v14i1.4743

Nabatchi, T., A. Sancino, and M. Sicilia. 2017. "Varieties of Participation in Public Services: The Who, When, and What of Coproduction." *Public Administration Review* 77 (5): 766–776. https://doi.org/10.1111/puar.12765.

Norman, D. 2013. *The Design of Everyday Things: Revised and Expanded Edition*. New York: Basic Books.

OECD. 2019. *Digital Innovation: Seizing Policy Opportunities*. Paris: OECD Publishing. https://doi.org/10.1787/a298dc87-en.

Office for Government Policy Coordination. n.d. "Government Performance Evaluation System." Retrieved January 17, 2025, from https://www.evaluation.go.kr/web/index.do

Olejniczak, K., S. Borkowska-Waszak, A. Domaradzka-Widła, and Y. Park. 2020. "Policy Labs: The Next Frontier of Policy Design and Evaluation?" *Policy & Politics* 48 (1): 89–110. https://doi.org/10.1332/030557319X15579230420108.

Osborne, S. P. 2018. "From Public Service-Dominant Logic to Public Service Logic: Are Public Service Organizations Capable of co-Production and Value co-Creation?" *Public Management Review* 20 (2): 225–231. https://doi.org/10.1080/14719037.2017.1350461.

Parasuraman, A., L. L. Berry, and V. A. Zeithaml. 1990. "Guidelines for Conducting Service Quality Research." *Marketing Research* 2 (4): 34–44.

Parasuraman, A., V. A. Zeithaml, and L. L. Berry. 1985. "A Conceptual Model of Service Quality and Its Implications for Future Research." *Journal of Marketing* 49 (4): 41–50. https://doi.org/10.1177/002224298504900403.

Parasuraman, A., V. A. Zeithaml, and L. L. Berry. 1988. "Servqual: A Multiple Item Scale for Measuring Consumer Perceptions of Service Quality." *Journal of Retailing* 64 (1): 12–40.

Service Design Tools. n.d. "Journey Map." Retrieved June 5, 2025, from https://servicedesigntools.org/tools/journey-map

Sundbo, J. 2015. "From Service Quality to Experience – and Back Again?" *International Journal of Quality and Service Sciences* 7 (1): 107–119. https://doi.org/10.1108/IJQSS-01-2015-0009.

Taplin, D. H., H. Clark, E. Collins, and D. C. Colby. 2013. Theory of Change - technical papers. ActKnowledge.

Tarena, E., P. London, A. Tangaere, P. Hagen, R. Taniwha-Paoo, and Kia Tipu Te Ao Mārama. 2021. "The Auckland Codesign Lab, The Southern Initiative." Tokona Te Raki *Kia Tipu Te Ao Mārama*.

The World Bank. 2018. *Indicators of Citizen-Centric Public Service Delivery*. Washington: World Bank Group. http://hdl.handle.net/10986/30030

Treasury Board of Canada Secretariat. 2021. "Theory-Based Approaches to Evaluation: Concepts and Practices." *Government of Canada*. https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/evaluation-government-canada/theory-based-approaches-evaluation-concepts-practices.html

Van de Walle, S. 2018. "Explaining Citizen Satisfaction and Dissatisfaction with Public Services." In *The Palgrave Handbook of Public Administration and Management in Europe*, edited by E. Ongaro & S. Van Thiel, 227–241. Palgrave Macmillan UK.

Van De Walle, S., and G. Bouckaert. 2003. "Public Service Performance and Trust in Government: The Problem of Causality." *International Journal of Public Administration* 26 (8-9): 891–913. https://doi.org/10.1081/PAD-120019352.

van der Bijl-Brouwer, M. 2016. "The Challenges of Human-Centred Design in a Public Sector Innovation Contex." Design Research Society Conference 2016.

van der Bijl-Brouwer, M., G. Kligyte, and T. Key. 2021. "A co-Evolutionary, Transdisciplinary Approach to Innovation in Complex Contexts: Improving University Well-Being, a Case Study." *She Ji: The Journal of Design, Economics, and Innovation* 7 (4): 565–588. https://doi.org/10.1016/j.sheji.2021.10.004.

Vargo, S. L., and R. F. Lusch. 2016. "Institutions and Axioms: An Extension and Update of Service-Dominant Logic." *Journal of the Academy of Marketing Science* 44 (1): 5–23. https://doi.org/10.1007/s11747-015-0456-3.

Vargo, S. L., M. A. Akaka, and C. M. Vaughan. 2017. "Conceptualizing Value: A Service-Ecosystem View." *Journal of Creating Value* 3 (2): 117–124. https://doi.org/10.1177/2394964317732861.

Welch, E. W., C. C. Hinnant, and M. J. Moon. 2004. "Linking Citizen Satisfaction with E-Government and Trust in Government." *Journal of Public Administration Research and Theory* 15 (3): 371–391. https://doi.org/10.1093/jopart/mui021.

Yang, S. B., and A. R. Torneo. 2016. "Government Performance Management and Evaluation in South Korea: History and Current Practices." *Public Performance & Management Review* 39 (2): 279–296. https://doi.org/10.1080/15309576.2015.1108767.

Yoon, S. W., J. W. Kim, N. Y. Ahn, I. A. Jung, C. Park, E. K. Ki, and J. S. Lee. 2014. *From Desk to Field: The Role and Application of Design in Public Policy and Public Services*. Rev. ed. Seoul: Korea Institute of Design Promotion & Robo & Company.

## Appendices

**Appendix A.** Delphi questionnaires - round 1 & round 2.
https://doi.org/10.6084/m9.figshare.28614641

**Appendix B.** An example of the relationships between codes and themes.
https://doi.org/10.6084/m9.figshare.28942238

**Appendix C.** Codes and themes list.
https://doi.org/10.6084/m9.figshare.28614683