International Review of Public Policy

8.3 | 2026
Special Issue

From play to policy success: The prospects of using serious games in public policies

Dominika Wojtowicz, Karol Olejniczak and Tomasz Kupiec

Electronic version
URL: https://journals.openedition.org/irpp/6632
DOI: 10.4000/16miz
ISSN: 2706-6274

Publisher
International Public Policy Association

Electronic reference
Dominika Wojtowicz, Karol Olejniczak and Tomasz Kupiec, "From play to policy success: The prospects of using serious games in public policies", *International Review of Public Policy* [Online], 8.3 | 2026, Online since 01 July 2026, connection on 29 July 2026. URL: http://journals.openedition.org/irpp/6632 ; DOI: https://doi.org/10.4000/16miz

From play to policy success: The prospects of using serious games in public policies

Dominika Wojtowicz, Karol Olejniczak and Tomasz Kupiec

AUTHOR'S NOTE

FUNDING INFORMATION:
This work was supported by the National Science Centre, Poland under Grant 2021/43/B/HS5/01935.

Introduction

1 Public policy practitioners, when developing and implementing public interventions, seek insights into what could work. In contemporary governance, this search increasingly unfolds under conditions of heightened turbulence, rapidly shifting contexts, compound crises, and accelerating feedback effects, which make it difficult to anticipate downstream consequences and to sustain core public values while adapting interventions over time (Ansell et al., 2024). Exploring the conditions for success or failure in a safe, non-threatening manner is especially challenging for public policies dealing with complex, non-linear problems – so-called wicked problems (Rittel & Webber, 1973; Head, 2019). These problems are inherently resistant to resolution, characterized by uncertainty, persistent value conflicts, dynamic contexts, and ambiguity in both problem definition and desired outcomes. In such settings, policymaking is shaped by interdependencies, path dependence, and nonlinear dynamics, which limit the usefulness of purely linear "best practice" transfer and call for approaches that support learning and adaptation (Cairney & Geyer, 2017). Moreover, value conflicts are not incidental but constitutive of many public

interventions; managing them is a core governance task that directly conditions legitimacy and perceived success (de Graaf et al., 2016).

2 Success or failure in such contexts can have multiple dimensions - programmatic, processual, and political (McConnell, 2010; Marsh & McConnell, 2010). Recent work on policy piloting further underscores that even "testing" interventions involves shifting purposes and multiple success criteria beyond program effects, including process and political dynamics (Checkland et al., 2023).

3 Given this, practitioners and governments now use experimentation and design-oriented approaches (such as pilots, regulatory sandboxes, and policy labs) to test interventions before full implementation. However, these approaches often require many resources, are hard to scale, and do not always connect with mainstream policy scholarship (Huitema et al., 2018; McGann et al., 2018; Olejniczak et al., 2020a). This situation creates demand for "safe-to-fail" methods. Such methods should help stakeholders make sense of situations and reveal trade-offs and constraints early. This is especially needed during turbulence and value conflict.

4 This demand is amplified by recent developments that reshape both practitioner expectations and how their decisions are judged. Digital transformation and the uptake of AI in government promise efficiency and responsiveness. However, they also intensify conflicts, such as transparency vs. privacy; efficiency vs. equity; automation vs. due process; and innovation vs. accountability. These tensions often arise in sensitive domains, including welfare, law enforcement, or migration governance (OECD, 2025). They are not just technical, they raise legitimacy questions about who sets the rules, how discretion is exercised, and how public authority is justified when decisions are mediated by complex socio-technical systems (Taylor, 2021). Even when institutions seek to implement "trustworthy" technologies, core values can remain in tension and, at times, be incommensurable. This makes procedurally fair deliberation and explicit trade-off management central to legitimate governance (de Fine Licht, 2025).

5 In this article, we argue that serious games (SG) have potential as a method for addressing and exploring policy success and failure for complex policy issues. SG is a game used for purposes other than entertainment (Abt, 1987; Flood et al., 2018). We conceptualize SG not as direct determinants of policy success, but as intermediate mechanisms that can support key policy-process capacities, such as joint sensemaking, coordination across actors, learning about behavioral and institutional constraints, and legitimacy-building thereby shaping the conditions under which programmatic, processual, and political success becomes more (or less) likely (Huitema et al., 2018; Olejniczak & Widawski, 2024). They are a highly versatile method used across different research areas and policy topics, and they have been dynamically developing in recent years using both analog means and cutting-edge technologies. In practice, their added value is that they allow multiple stakeholders to "rehearse" decisions under constraints, surface hidden assumptions, and experience value trade-offs, often in participatory formats that can also strengthen engagement and civic learning (Hassan, 2020). As a complement to sandboxes and other experimentation infrastructures, SG can offer a comparatively low-cost and scalable way to explore policy options, stress-test implementation logics, and structure deliberation before high-stakes commitments are made (Furber et al., 2018; World Bank, 2020).

6 However, leading public policy journals and mainstream literature rarely address the role of games in policy success and failure, creating a "diffusion gap" between the growing body of SG research and the wider policy world. This gap limits ongoing learning about how games can shape policy design, implementation, and evaluation. By situating SG in broader debates on experimentation, learning, and legitimacy in complex settings (Huitema et al., 2018; McGann et al., 2018; de Graaf et al., 2016), we show their potential as a practical, simulation-based complement to current approaches. This is a knowledge gap and an opportunity worth exploring.

7 Therefore, in this article, we aim to address three questions:

> Q1: What is the actual trend in the use of SG in public policy literature?
>
> Q2: What are the preferences of policy practitioners towards SG?
>
> Q3: What are the prospects of using SG in the area of policy success?

8 The discussion in this article is based on the empirical materials collected during three literature reviews on SG in public policy and multiple consecutive surveys with policy practitioners (for details, see Section 2). By combining a structured mapping of the literature with practitioner-oriented evidence on perceived usefulness, we aim to clarify both the state of the field and the practical feasibility conditions for wider uptake.

9 The article's structure is as follows. In the next section, we provide a detailed definition of SG and present a typology that groups a wide spectrum of SG into four basic groups: teaching games, assessment games, Crash-testing Mechanisms games, and exploration games. This provides a conceptual frame for our empirical exploration and discussion. This is followed by the Method section. In the third, Empirical section, we show the growing popularity of SG in scientific literature and contrast this trend with the still limited discussion in public policy and public administration journals. We follow this with survey results indicating policy practitioners' preferences for specific types of SG. In the discussion section, we explore how four types of SG could be useful for enhancing the specific dimensions of policy success identified by the seminal works of Marsh and McConnell (2010) and McConnell (2010), namely process, programmatic, and political success. We conclude by summarizing the opportunities for applying SG in public policy and suggesting directions for future research and practice.

10 Our primary contribution is to position SG as a method that advances policy design and evaluation by operationalizing established frameworks of policy success. We show that aligning types of SG with the process, programmatic, and political dimensions expands the utility of games beyond conventional educational uses. This typology clarifies how simulation-based methods reveal policy contingencies and trade-offs, especially in complex environments. Thus, we connect theories of wicked problems to practical methodologies for addressing them.

11 For practitioners, this article provides a framework for choosing SG tailored to policy needs. Using both literature review and practitioner feedback, we show how SG can help explore policy ideas in low-risk, reflective settings. These findings can support innovation units, policy labs, and public agencies that want more participation, robust policies, and tested assumptions. As governments expand experimentation, our

typology can guide design, learning goals, and simulation choices (McGann et al., 2018; Olejniczak et al., 2020b). Using new technologies like generative AI for game analysis, we offer a scalable way to bring SG into real policy work.

Conceptual section

Origin and definition of SG

12 Historically, games were applied for purposes beyond entertainment, notably in educational and military contexts. In particular, war games from the 17th and 18th centuries paved the way for simulation-based learning, which would significantly influence SG (Wilkinson, 2016).

13 In modern public policy context, the term "serious game" was formally introduced by Clark Abt with his 1969 seminal publication, and defined as "*games that have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement [...] although it does not mean that they are not or should not be entertaining*" (Abt, 1987, p. 9). Significantly, Abt's conception covered both digital and non-digital formats, reflecting his experience developing military training simulations. Richard D. Duke further extended the SG applications as a method to improve communication between competing policy stakeholders (Duke, 2011, p. 343).

14 The "current wave" of SG began around 2002, largely popularized by the Sawyer & Rejeski (2002) white paper and the subsequent Serious Games Initiative that advocated for leveraging commercial video game technology and resources for public organizations. Contemporary technology opportunities and the growth of the commercial video game industry have increasingly narrowed the focus to SG definitions and practices primarily to digital applications (Connolly et al., 2012; Çiftci, 2018; Maxim & Arnedo-Moreno, 2025).

15 In this article, we follow the extended definition of SG recently adapted to the field of public policy: *"Serious games (SG) are analog or digital games that have a clear primary purpose different from entertainment, and that intentionally link/transfer the game experience to reality. They also have clear rules, and are used within a well-defined space (e.g., classroom, training program, research project). Both the purpose and a way of transfer are determined by the game principal, who is a person or organization that applies the game to a public policy issue."* (Olejniczak et al., 2020b, p. 345).

16 This definition encompasses both digital and analog forms. There are three two reasons for this. First, it follows the original, broad conceptualization of the first promoters of the SG for policy (Abt, 1987; Duke, 2011). Furthermore, this framing allows us to examine analog games that are easier and quicker to develop and that facilitate more traditional face-to-face interactions in decision rooms, which are particularly beneficial for policymaking processes.

Typology of SG for public policy evaluation

17 Game designers have developed several typologies of games. (Elias et al., 2012). However, only a few of these conceptual frames take into account the specificity of public policy (Duke & Geurts, 2004; Mayer et al., 2016). They are often conceptually

intricate, limiting their practical applicability. Given this scarcity of straightforward frameworks tailored to public policy, we rely on the evaluation-based classification proposed by Olejniczak, Newcomer, and Meijer (2020b) as the most relevant and pragmatic model. Although it originates from an evaluation perspective, this framework offers a clear structure that we adapt to explore the broader use of SG across multiple stages and dimensions of public policy.

18 This typology rests on two dimensions. The first dimension concerns the game's primary purpose: either testing change or triggering it. Some games act as tools enabling controlled observation of behaviors or system dynamics in response to policies, functioning like simulations that generate data on reactions under defined conditions. Others function as interventions themselves, designed to build capacities, foster shared understanding, and promote behavior change among participants. The second dimension relates to the evaluand (what is being assessed), distinguishing between complicated systems, which are characterized by stable processes and predictable relationships (e.g., regulatory procedures), and complex systems, which are marked by emergent behaviors, non-linear dynamics, and multiple actors that are difficult to predict or control (e.g., community resilience or environmental transitions). While this typology is rooted in evaluation contexts, for the purposes of our analysis the evaluand will correspond broadly to characteristics of public policy issues.

19 Combining these dimensions produces a two-by-two matrix that identifies four typical areas of SG application in public policies design and implementation process.

20 (1) **Teaching Knowables** are games that help participants learn accepted principles within structured systems. These games build understanding of well-established concepts and procedural knowledge in stable, predictable systems. Often used to train policymakers and officials, they enable learning of effective strategies and theoretical models in environments with established "right answers." These games serve as instruments of policy learning, enhancing practitioners' understanding of public governance mechanisms. For example, they help users grasp behavioral insights, collective action dynamics, or institutional design principles, contributing to more informed policymaking.

21 (2) **Testing Retention** games assess players' real competencies and behaviors by observing performance in controlled yet dynamic environments. Unlike traditional testing methods, these games assess applied competencies: how people act under pressure, constraints, or uncertainty. The aim is to observe whether participants can apply their training and knowledge in practice. These games help identify skill gaps, validate implementation readiness, and inform resource allocation decisions. They are especially useful in structured systems with known procedures, where the question is whether actors can implement them effectively. Their utility depends on whether the game reflects real challenges and measures competencies needed for real-world performance.

22 (3) **Crash-testing Mechanisms** simulate complex systems to explore how public interventions trigger outcomes or affect existing systems. These games simulate policy environments to test how legal, institutional, procedural, or infrastructural systems might react under disruptive conditions. They assess potential effects of policy interventions and stress-test systems against hypothetical events, crises, or systemic shocks (e.g., natural disasters, large-scale migration, pandemic outbreaks). They help identify vulnerabilities, emergent patterns, unintended consequences, and policy blind

spots. Crash-testing Mechanisms games are especially relevant when evaluating existing arrangements under uncertainty or rapid change.

23 (4) **Exploring Systems** games address complex policy challenges characterized by uncertainty, value pluralism, and competing interests. Unlike Crash-testing Mechanisms games that test specific mechanisms, these games facilitate collective sensemaking around "wicked problems" that resist clear solutions. They create environments where stakeholders can explore systemic interdependencies, experiment with policy alternatives, and build shared understanding of challenges. These games are distinguished by their open-ended nature and emphasis on collaborative problem-solving. Challenges are deliberately vague, with outcomes emerging through dialogue rather than predetermined rules. Game facilitators guide processes while players become co-producers of solutions through collective exploration. The games' outcomes include enhanced empathy, increased social capital, and improved capacity for navigating competing interests. They prove valuable in domains such as climate change, social integration, and territorial governance, where technical solutions alone are insufficient and political dimensions require equal attention.

Table 1. Typology of serious games.

<table>
<tr><td></td><td colspan="2">COMPLEX ISSUE</td><td></td></tr>
<tr><td rowspan="2">TESTING CHANGE</td><td>CRASH-TESTING
Assess possible system's reaction to intervention</td><td>EXPLORING SYSTEMS
Learn other actors' perspectives and system's dynamic</td><td rowspan="2">TRIGGERING CHANGE</td></tr>
<tr><td>TESTING RETENTION
Assess people's real, not declarative competences</td><td>TEACHING KNOWABLES
Learn principles, concepts and effective strategies</td></tr>
<tr><td></td><td colspan="2">COMPLICATED ISSUE</td><td></td></tr>
</table>

Source: adopted from Olejniczak et al., 2020b

24 The typology assumes that SG are not limited to one-off experiments or training tools but may represent a versatile instrument capable of serving a broad spectrum of purposes – from generating evidence to fostering shared values or understanding – across both structured and fluid policy contexts. In the discussion section, we will illustrate each category with examples drawn from practice.

Methods

25 Our research design comprised three interconnected streams of activities. First, we conducted three literature reviews over consecutive years to examine the use of SG in public policy. We used structured search queries across databases and manual journal scans. Second, we carried out multiple practitioner surveys in diverse settings to gather policy professionals' perspectives on SG, ensuring anonymity and ethical compliance. Third, we identified diverse illustrative examples of SG used in public policy within the final corpus of publications retrieved and screened during the reviews and used AI-

assisted tools to support extraction and coding, with human verification of all outputs. Importantly, this third stream is not a standalone search for cases, but a within-corpus identification and extraction step.

Literature search, bibliometric mapping, and case identification

26 Literature reviews were conducted over three consecutive years and focused on the applications of SG in public policy. In the first two searches (University of Washington, Seattle: September–December 2023; Temple University, Philadelphia: June–August 2024), we used the Web of Science database and a simple search query TS=("serious gam*") AND TS=("public polic*").We also manually reviewed publications from the last 10 years in two key SG journals (*Simulation and Gaming*, *International Journal of Serious Games*) and selected public policy journals (*Policy Sciences*, *International Review of Administrative Sciences*, *Public Administration Review*, *Policy & Society, American Journal of Evaluation*). This was accompanied by a search for books on the intersection of games and public policy (Amazon US store and major publishers on the topic, such as Springer and MIT), used primarily to strengthen conceptual framing and to cross-check terminology rather than to quantify publication trends. The outputs of these first two policy-focused searches (2023 and 2024) formed the review corpus from which we subsequently identified and extracted illustrative SG examples used in the Discussion section where we examine how different SG families may contribute to policy success. The third search (Kozminski University, Warsaw - June-July 2025) served a bibliometric mapping purpose (trend analysis). It was performed systematically through the Web of Science database (search formula: TS=("serious gam*"), all languages, all document types, 1900–2025). This broad mapping corpus was used to derive the aggregate publication trends and disciplinary distribution reported in Section 3. Additionally selected public policy journals were scanned with a specific search formula (search on 29 July 2025 with search query: TS=("serious gam*") AND SO=("Public Administration Review" OR "Governance" OR "Policy Sciences" OR "Policy and Society" OR "Public Policy and Administration" OR "Policy & Politics" OR "Journal of Comparative Policy Analysis" OR "Policy Design and Practice" OR "Behavioral Public Policy" OR "Evidence & Policy")). The Web of Science was chosen for its coverage of both social and hard sciences.

27 Identification of illustrative examples within the review corpus (from the 2023–2024 policy-focused searches). Based on the final set of publications included after screening in the policy-focused searches, we conducted a structured within-corpus scan aimed at identifying diverse illustrative examples of SG used in public policy across domains. Accordingly, this step should not be interpreted as a comprehensive inventory of all SG used in public policy, but as an example-driven selection strategy grounded in the reviewed literature. From this corpus, we retained 20 illustrative examples used in the Discussion section. The analysis of collected sources was conducted in two ways: book content was analyzed by a human researcher. In contrast, article content was analyzed by researchers supported by selected generative AI agents. AI tools were introduced only after candidate examples had been located in the included publications, and all extracted information was verified against the full texts.

28 The table below provides details on tasks for which we applied AI tools, ways of validating their output as well as human oversight.

Table 2. The table below summarizes the tasks performed by the AI tools, the validation of AI outputs, and the human oversight to ensure data integrity.

Task	AI agent	AI output	Validation method	Human oversight
Task 1: Quick overview of articles' content	Petal AI - table" option (OpenAI GPT-4 via RAG)	Comparative table with 2-paragraph summaries per prompt, answering 5 structured questions: (i) definition of SG, (ii) game name, (iii) game purpose, (iv) target players, (v) game design and mechanics	No direct text-to-summary links available; outputs required manual cross-referencing with source articles	Human analysts compared AI-generated notes against full article content for accuracy verification
Task 2: In-depth exploration of individual articles	MAXQDA 24 AI Assist – „chat with documents" (GPT engine)	Extended responses with references to specific paragraphs, covering: game purpose, form (digital/analog/hybrid), recorded effects and impact, and game type classification	Direct traceability between AI summaries and underlying evidence through paragraph-level references	Researchers verified AI-generated paragraph references against source documents and assessed classification accuracy
Task 3: Pros and cons analysis of case utility for policy success	ChatGPT and Claude AI - dual-agent approach (GPT and Claude Sonnet)	Arguments for and against how specific game types could support each of the three dimensions of policy success, based on provided game type descriptions and policy success definitions	Cross-validation through comparison of outputs from two independent AI systems	Human researcher evaluated and synthesized arguments from both AI agents, assessing coherence and relevance

Source: own elaboration.

29 Using generative AI to support extraction, coding, and brainstorming carries methodological costs that shape how our findings should be read. Three risks are particularly relevant for the way these tools were deployed in this study, and our design contains specific safeguards against each of them.

30 The first is related to hallucination. Large language models can produce fluent statements that are not anchored in the underlying source, and this risk does not disappear when a model is coupled with retrieval-augmented generation, as in the case of Petal AI used in Task 1 (Reinhard et al., 2025). For this reason, we did not rely on the comparative summaries on their own. Each Petal AI output was verified by a human analyst against the full text of the source article before any claim was included in the synthesis. Task 2 was designed precisely to address the traceability gap of Task 1: MAXQDA's chat-with-documents function anchors responses to specific paragraphs of

the analyzed articles, which allowed us to trace each interpretive claim back to a locatable passage.

31 The second risk concerns reliability and replicability. Outputs of contemporary LLMs are sensitive to prompt phrasing, model version, and stochastic sampling, which limits exact reproducibility of any AI-assisted extraction or coding step (Thomas et al., 2026; De Paoli, 2024). Recent systematic evidence on generative AI in evidence synthesis emphasizes the same caution: AI assistance can accelerate screening and extraction, but accuracy varies across tasks and corpora, and human verification remains indispensable (Clark et al., 2025). Accordingly, we treated AI outputs as drafts to be checked, not as findings, and we used identical prompts across articles to keep extraction conditions comparable.

32 The third risk relates to AI used as a brainstorming partner. In Task 3 we used ChatGPT and Claude to generate pros and cons for how SG types could support each dimension of policy success. Experimental work shows that dyadic human-LLM brainstorming can improve fluency of ideas but can also induce cognitive anchoring on AI-proposed framings (Memmert et al., 2025). Two design choices reduce that risk in our case. We triangulated outputs of two different model families rather than relying on a single model, and we retained only those arguments that could be grounded in the reviewed evidence. We cannot rule out that the framing of the Discussion section has been subtly shaped by structures the models found most plausible, but the brainstorming step did not introduce claims that did not survive evidence-based filtering.

33 All these caveats bound the epistemic weight of the AI-supported steps in our analysis without invalidating them. The illustrative SG examples and the typology mappings developed in the Discussion section rest on AI-assisted extractions that were verified against full texts and on brainstorming outputs that were filtered against reviewed evidence.

Survey with practitioners

34 The survey was conducted on multiple occasions across different settings between 2020 and 2023, each involving distinct groups of policy practitioners. Data collection took place during academic lectures at universities in the United States and Poland specializing in public policy, community-of-practice sessions with employees of a leading international development organization, workshops at an international evaluation conference, and specialized policy training programs. The total sample comprised 143 respondents.

35 The survey was designed as an exploratory mapping exercise rather than as a representative study of the policy profession. Its primary purpose was to identify patterns in perceived utility of different serious-game types across practitioner groups and institutional contexts. Accordingly, the findings are interpreted as indicative signals that inform the conceptual argument of the article, rather than as statistically generalizable estimates of practitioner attitudes.

36 Given the nature of the study (an anonymous, low-risk survey collecting only opinions on game preferences without any personally identifiable information or sensitive data), formal ethics approval was not required under the institutional guidelines applicable to

these settings. The study posed no foreseeable risks to participants beyond those encountered in everyday professional discussions.

37 Informed consent was obtained through a transparent disclosure process. Prior to survey administration, participants received a verbal explanation that the data collection was part of a research exercise aimed at understanding policy practitioners' perspectives and bridging the gap between game designers and policy practitioners. Participation was entirely voluntary, and respondents could choose to stop the survey at any time with no consequences.

38 Data protection was ensured through the instrument's anonymous design. The survey collected no demographic information or identifying characteristics: only preferences regarding four types of SG rated on a five-point scale (1 = low utility, 5 = high utility). Depending on the context, the survey was administered either online via mobile devices or in paper-based format. In both cases, responses could not be linked to individual participants, ensuring complete anonymity.

39 The sample was purposive and context-dependent. It consisted of participants in academic programs, professional workshops, and evaluation-oriented settings. The sample does not constitute a representative cross-section of public officials across administrative tiers, sectors, or policy domains. External validity is therefore limited, and results should be interpreted as exploratory patterns rather than population-level inferences.

40 Survey results are presented using descriptive statistics, while differences between groups were tested using linear regression models. In these models, the types of games served as dependent variables (average score on a 1–5 scale), and professional status (current practitioner = 1, future practitioner = 0) and the maturity of the evaluation culture in the respondent's country (mature = 1, non-mature = 0) were included as independent variables. Interaction effects between the two independent variables were also examined. Statistical significance was assessed at the $\alpha = 0.05$ level. Cluster-robust standard errors were not applied due to the small size of clusters (limited sample sizes within specific time–location combinations).

41 Although the underlying items were measured on a five-point Likert scale, the averaged scores were treated as approximately continuous variables. Treating Likert-type scales with five or more ordered categories as interval-level for parametric analysis is common practice in applied social science research, particularly when composite or mean scores are analyzed (Carifio & Perla, 2008; Norman, 2010). Simulation and robustness studies show that parametric methods such as linear regression are generally robust to moderate violations of the interval assumption in Likert-type data, especially in samples of this size (Norman, 2010). Given the exploratory purpose of the survey and the focus on directional group differences rather than precise population estimates, linear regression was selected for interpretability and parsimony.

Empirical evidence: bibliometric patterns and practitioner survey

Trends and patterns of publications on SG

42 In recent years, SG have gained increasing attention across a wide range of scientific disciplines, application domains, and fields, reflecting their versatility and adaptability to various research and practical purposes. The growing body of academic literature on SG demonstrates this trend. A bibliometric analysis conducted using the Web of Science Core Collection identified 11,280 publications addressing the topic.[1]

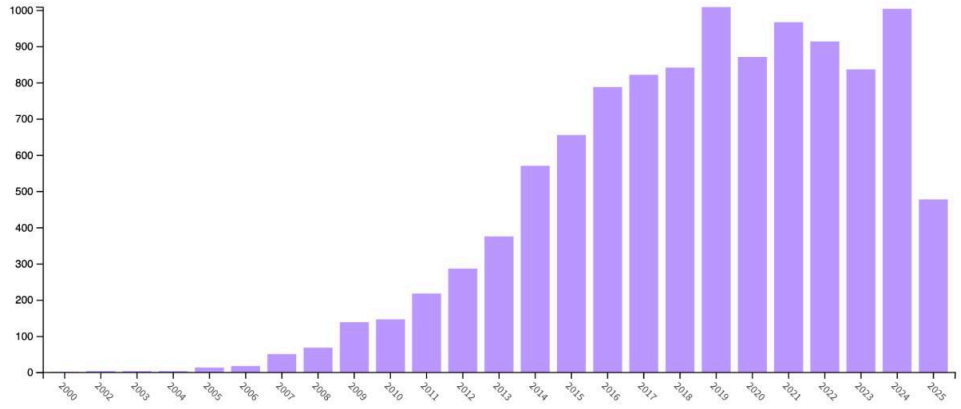
43 The disciplinary distribution of this research is diverse. As visualized in Figure 1, the largest clusters of publications are located within Educational Research (2,576 records) and Computer Science Interdisciplinary Applications (2,168 records), followed by Computer Science Theory Methods (1,732 records) and Computer Science Information Systems (1,380 records). Other significant fields include Software Engineering, Artificial Intelligence, Health Care Sciences, and Medical Informatics. This disciplinary landscape highlights the multidisciplinary character of SG research, which spans both technical and social sciences domains.

Figure 1. Disciplinary distribution of SG publications (n=11,280).

Source: WoS 29.07.2025

44 The temporal distribution of publications (Figure 2) shows a steady increase in interest since the mid-2000s, with the number of publications peaking between 2014 and 2023, consistently exceeding 800 records per year. Between 2000 and 2025, the number of publications related to SG increased fivefold. Compared to the overall increase in all publications in the WoS database, the growth in articles on this topic was more than twice as high. This trend indicates that SG are not only an established area of academic interest but also a continuously evolving research field.

Figure 2. Temporal distribution of SG publications (2000–2025).

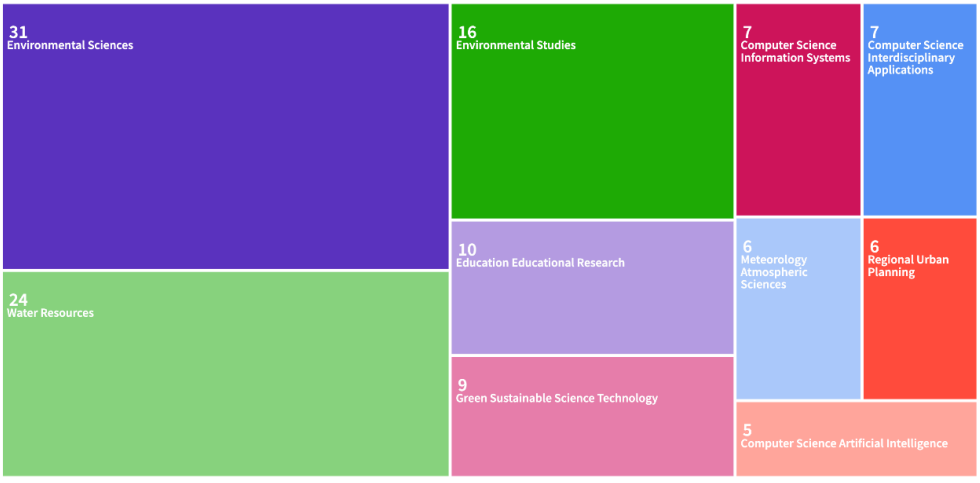


Source: WoS 29.07.2025

45 When the corpus is narrowed to publications explicitly addressing SG within the context of public policies — using the query TS=("serious gam*") AND TS=("public policy" OR "public policies" OR "policy-making" OR "governance") — a total of 113 publications were identified. These are distributed across multiple research areas as illustrated in the graph (Figure 3).

46 The majority of these publications are concentrated in environmentally oriented disciplines, notably *Environmental Sciences* (31 publications), *Water Resources* (24), and *Environmental Studies* (16). SG also appear in educational research (10 publications), particularly in the context of science communication and stakeholder engagement. There is also a visible representation in computer science disciplines, including *Information Systems* (7), *Interdisciplinary Applications* (7), and *Artificial Intelligence* (5).

47 Figure 3. Disciplinary distribution of SG in public policy (n=113).



Source: WoS. 29.07.2025

48 However, the topic of SG is virtually absent from leading public policy journals. We identified only three articles that reference SG in their title, keywords, or abstract. The earliest of these was published in 2018, with two appearing in Policy Design and Practice and one in Behavioral Public Policy (Olejniczak et al., 2018; de Smedt & Borch,

2022; Krpan & Urbaník, 2024). This contrast points to a diffusion gap between communities: SG research expands rapidly across adjacent fields, yet the topic enters mainstream policy venues only sporadically. Rather than indicating a lack of applications, the pattern suggests limited translation into the publication outlets that structure cumulative debate in public policy and public administration.

Preferences of policy practitioners – survey results

49 The perception of policy practitioners on the potential utility of SG for public policy was explored with the survey. These were academic course participants from US and Polish universities specialized in public policy, employees of a leading international development organization, participants of an international evaluation conference, and attendees of specialized policy training programs. The total sample comprised 143 respondents, with a relatively balanced distribution across two key characteristics: the respondents' status as policy practitioners and the maturity of the evaluation culture within their national administrations (see Table 3).

Table 3. Structure of the sample

		Evaluation culture	
		Mature	Non-mature
Practitioners	Current	34	33
	Future	43	33

Source: own elaboration.

50 Surveyed individuals were asked to assess the potential utility of four types of SG for their work, on a five-point scale with 1 described as low and 5 described as high.

51 All four types of games received moderately positive responses from our respondents, with average scores above 3 on a scale from 1 to 5. The highest-rated application of games was their potential for exploring systems, that is, building understanding of the complex system dynamics underpinning particular policy problems (3.82)[2]. The lowest-rated application was testing retention, understood as the assessment of knowledge, skills, and competencies among stakeholders involved in a specific policy or program (3.08).

Table 4. Mean scores for game types.

Teaching knowables	Testing retention	Crash-testing Mechanism	Exploring systems
3,63	3,08	3,50	3,82

Source: own elaboration.

52 Differences in perceptions of game types were also observed depending on respondents' professional status and the maturity of the evaluation culture in their country.

53 Current practitioners rated Crash-testing Mechanisms games on average 0.58 points higher than future practitioners. However, this effect was observed only in countries with less developed evaluation cultures, where the difference reached 0.80. In contrast, in countries with more mature evaluation cultures, no significant difference was found between current and future practitioners.

54 For retention-testing games, current practitioners gave consistently higher ratings than future ones, with an average difference of 0.40. In this case, the level of evaluation culture had no significant moderating effect.

55 The pattern was more complex in the case of system-exploring games. Among current practitioners, those from countries with more mature evaluation cultures rated these games slightly lower (−0.26) than their counterparts from countries with less developed evaluation cultures. However, among future practitioners, the pattern was reversed: those from more developed cultures rated system-exploring games significantly higher than their counterparts from less developed ones (+0.56). In less developed evaluation cultures, current practitioners rated system-exploring games more positively (+0.48), while future practitioners rated them more positively in more mature cultures(+0.34).

56 Overall, the reception of SG among respondents, both current and future policy practitioners, was cautiously optimistic. All game types received moderately positive ratings, with the highest scores associated with their potential to explore complex systems and more generally facilitate change. Notably, these assessments were made following only a brief introductory lecture, without participants having played the games themselves, let alone having used them in professional practice. These patterns suggest that even limited exposure to the concept of SG can generate interest and perceived relevance within the policy community. Moreover, the consistently higher ratings provided by current practitioners may carry particular weight, as their direct experience with policy design and implementation provides a more informed perspective from which to assess the potential value of SG as innovative, experience-based tools. These findings may point to a favorable initial orientation toward SG in policy evaluation and justify further piloting, testing, and capacity-building efforts.

57 These findings should be interpreted as exploratory signals rather than definitive evidence. The survey captures respondents' assessments following brief conceptual exposure to serious games, rather than after sustained institutional use. As such, the results illuminate the perceived plausibility and relevance of SG across practitioner groups, while questions of effectiveness, uptake, and long-term impact remain for future, larger-scale research.

Discussion: prospects of using SG for policy success

58 In the beginning of this article, we argued that when integrated into public policy processes, SG can contribute to policy success and help reduce the risk of failure. However, our review also suggests that uptake among public policy practitioners remains cautious and that the use of SG is still uneven across policy fields and settings.

Against this backdrop, this section examines the prospects and conditions under which the four types of SG discussed in this article can increase the likelihood of successful policy outcomes and support the anticipation and avoidance of policy failures.

59 We frame this discussion using the conceptualization of policy success developed by McConnell (2010) and Marsh and McConnell (2010), who distinguish three interrelated dimensions: process success, programmatic success, and political success. In what follows, we use this three-dimensional lens as our analytical point of reference and apply it to the broader institutional space in which public interventions are designed and delivered rather than solely to the government tier. Our vantage point remains policymaker-centric, understood broadly as officials responsible for the planning, implementation, and/or evaluation of public interventions (regulations, projects, programs, or long-term policies) across tiers and sectors. We then consider how the four SG types introduced in the previous section, i.e. Teaching Knowables, Testing Retention, Crash-testing Mechanisms, and Exploring Systems, can be used for distinct policy tasks and decision needs that arise across the policy process.

60 While we structure the analysis using the three-dimensional lens of policy success, we also acknowledge the recurring mechanisms, understood as pathways through which an effect is produced, that each type of SG may activate: coordination, learning, and legitimacy (Gerring, 2007). Games designed for Exploring Systems center on building a shared representation of a complex problem through collective exploration of scenarios and trade-offs; this makes coordination (collective sensemaking and alignment across actors) a particularly plausible pathway, consistent with work on boundary objects and boundary crossing as enabling coordination across communities and perspectives (Star & Griesemer, 1989; Akkerman & Bakker, 2011). At the same time, such system exploration typically entails learning as participants observe feedback loops and consequences across scenarios (Geurts et al., 2007), and it can also support legitimacy when deliberation is structured in ways that make assumptions and trade-offs transparent and inclusive (Fung, 2006; Tyler, 2006). Crash-testing Mechanisms games are designed to probe assumptions, stress-test policy designs, and surface failure points; accordingly, they primarily operate through learning driven by feedback and diagnosis – often complemented by coordination around shared assumptions and roles – consistent with experiential learning and instructional game models (Kolb, 1984; Garris et al., 2002). Teaching Knowables games target domains where knowledge and procedures can be trained and standardized; their dominant pathway is likewise learning (skill acquisition and procedural repertoires), again consistent with learning-by-doing and reflection logics (Kolb, 1984; Garris et al., 2002). Finally, Testing Retention games assess whether participants can reliably apply known procedures under constraints and uncertainty; their dominant pathway is learning (retention) in the sense of identifying competence gaps and strengthening operational readiness through repeated practice and feedback (Kolb, 1984; Garris et al., 2002).

SG and process success

61 Process success concerns the legitimacy and robustness of policy formation: how issues are framed, options explored, interests heard, and decisions justified through procedures that sustain due process and accountability. In practice, this dimension is often judged based on the perceived fairness and transparency of deliberation, the

ability to reach decisions without hollowing out core aims, and the durability of support behind an adopted course of action. Within this dimension, SG matters mainly through a legitimacy pathway: they can structure deliberation, make trade-offs and assumptions explicit, and leave a traceable record of how disagreement was handled and compromise formed. Used in this way, SG can complement (rather than replace) conventional consultation formats.

62 **Exploring Systems** SG are especially well-suited to this purpose. By simulating complex interdependencies, including budgetary and legal constraints, time lags, and side-effects, they let participants experience real-world complexity rather than merely comment on abstractions. Such SG make feedback loops, thresholds, and rebound effects visible and allow players to adjust constraints to observe how the system responds. They also support procedural fairness (clear rules, balanced voice) and perspective-taking (roles and shared models), helping participants converge on feasible policy bundles that combine measures, sequencing, and acceptance conditions.

63 In this vein, the "Consensus Building Game" (Oyamada & Kitani, 2019) illustrates how such formats can build compromise around policy bundles in rural policy contexts. The board-and-card format stages a human–wildlife dilemma: crop damage by wild boar versus ethical and ecological arguments for conservation. Participants take grounded roles (e.g., farmers, local residents) and move through three linked phases. Across gameplay, dilemma presentation, and remodeling, participants surface value conflicts and iteratively redesign rule bundles (e.g., culling thresholds, fencing/compensation mixes, seasonal sequencing) until they reach explicit acceptance conditions. The process yields usable artifacts such as revised rules, bundled measures, and short rationales that document how compromise formed. These artifacts can be summarized in a consultation record.

64 The use of SG as a public-consultation tool offers another benefit: an auditable deliberation trail demonstrating inclusiveness and due consideration to the policy issue under consultation. "SPRINTCITY" (Duffhues et al., 2014), a computer-supported, multiplayer planning game, shows how such a trail emerges in practice. Municipalities, a regional authority, and the rail operator take interdependent turns to set station-area land use, service levels, and investment phasing under policy and budget constraints. After each round, the shared spatial model updates transparent indicators (e.g., accessibility, capture of corridor demand, train occupancy, costs), anchoring discussion in a common evidence base. Facilitated debriefs translate disagreements into documentable outputs (scenario comparisons, prioritized project sets, phasing sketches, concise rationales), while the platform's records (logs and exported model outputs) create a traceable deliberation trail that can accompany consultation summaries and support an impact assessment.

65 Finally, SG can widen inclusion by engaging groups who would not attend traditional consultations due to sensitivity of the topic, logistical barriers, or lower procedural literacy. "The People" an analogue, story-driven simulation of ethically charged dilemmas around managing illegal crossings at the Polish–Belarusian border, demonstrates this potential (Szatkowska et al., 2025). Players inhabit roles of local residents within a government-restricted zone; hidden motives create tensions among neighbors, while the board tracks the fate of refugee groups and reveals real, documented stories only if a group is 'saved.' The narrative framing lowers the threshold for participation, and small, facilitated groups make the experience

accessible. In the debrief, participants articulate priorities and conditions for support (e.g., acceptable safeguards, non-negotiables, preferred sequencing of measures) and translate insights into concrete proposals (adding a safeguard, adjusting thresholds, reordering actions). The outputs, i.e., role-based notes, agreed-upon criteria, and brief rationales, become process artifacts that can be transcribed into a participation record. These formats can elicit views that are difficult to capture in conventional meetings. Through the use of plain language, multilingual materials, and accessibility accommodations, participation can be expanded beyond the "usual voices."

SG and programmatic success

66 The second dimension of policy success is programmatic success, i.e., policy performance in practice, gauged by whether a measure is implemented as approved, achieves its intended effects, and does so with prudent use of resources. Programmatic success is typically assessed in terms of implementation fidelity, outcome attainment, efficiency, and distributional fit across target groups and stakeholders; evidence may come from evaluations, audits, stakeholder reviews, and professional/media scrutiny. Assessments are often mixed across criteria rather than binary.

67 This is the dimension where SG have the broadest applicability. Used across formulation, delivery, and evaluation, each of the four SG families can, in complementary ways, increase the likelihood of programmatic success, especially by bolstering the attainment of intended effects, while also supporting operational fidelity, efficient resourcing, and fit with target groups' needs.

68 The first game type, i.e., **Teaching Knowables** can raise the knowledge levels of both policy-makers and policy-takers in ways that support more effective delivery on policy objectives and measurable results.

69 Turning first to policy-takers, Teaching Knowables games serve as a practical capacity-building tool for policy implementers - from central ministries and agencies to regional authorities, local government departments, and frontline managers. By encoding accepted principles and rule-bound procedures, they help officials build shared vocabulary and rehearse standard decision rules before applying them in real programs.

70 A good example is "Healthene™", a board game designed on value-based healthcare financing (Tan et al., 2020). Players, who are finance officers in the public health system, move patients through a system of acute/community/home care, while deploying Funding Innovation cards (e.g., preventive and chronic care, bundled payments, frailty bundles, end-of-life plans) to see how recognized financing tools affect costs and pathways. Transparent feedback supports practice in allocation and costing logic relevant to implementation. Another example is the "Groundwater Management Game", which was developed in the context of implementing California's Sustainable Groundwater Management Act (Goodspeed et al., 2020). The aim of the game was to use an accessible simulation to teach public officers (including members of local GSAs, water-system utility staff, and state or federal officials) the resource-management tools authorized under the Act (e.g., pumping limits, allocations, simple mechanisms for trading entitlements) and make public officers aware of their consequences for the shared groundwater aquifer.

71 When addressed to policy takers: stakeholders and target groups, Teaching Knowables SG can be used for two complementary purposes: (1) providing procedural instruction for the rollout of public interventions and (2) building system understanding and readiness for behavior change.

72 In the first case, games may serve as practical instructional tools when a new or complex intervention is being introduced. They let those most affected rehearse procedures and decision rules in a safe, low-stakes environment, while feedback helps surface misunderstandings before implementation. This reduces avoidable implementation errors and improves fidelity during rollout. One example of such a game is "Changing the Game – Neighbourhood", a board game that operationalizes established energy-transition rules so residents, community groups, local officials, and utility partners can practice them before real-world rollout (Lanezki et al., 2020). Players cooperatively set targets for electricity, heating, and mobility (e.g., CO_2 and energy-savings goals) and then combine standard measures and technologies, such as PV, storage, heat pumps, building insulation, mobility options, or demand shifting by playing cards that represent recognized options and procedures. The feedback loop is immediate and transparent: participants receive instant information on energy flows, CO_2 intensity, and costs for each option they select. Across rounds, participants rehearse rule-based choices and can translate outputs (targets, selected measures, rationales) into follow-up guidance.

73 In the second case, Teaching Knowables SG can build stakeholders' and target groups' knowledge aligned with policy goals and support readiness to adjust behaviors. For instance, "The Plastic Pipeline" is an interactive online game in which participants act as detectives investigating the life cycle of single-use plastic waste; it clarifies environmental risks and the logic of policies such as bans or levies and makes visible links among producers, consumers, and regulation (Vecchio & Del Greco, 2023). Similarly, "Recovery Rapids" supports learning concrete pro-environmental behaviors through immediate feedback (Fox et al., 2020): the player controls a kayaker cleaning a polluted river and sees real-time consequences of actions (e.g., waste segregation, pollution prevention). In the health-policy domain, "Fresh Food Runner" (Petko et al., 2020) uses an arcade-style runner format where players collect seasonal fruits and vegetables while avoiding off-season or ultra-processed items; short prompts and end-of-level summaries translate choices into simple indicators (e.g., health and sustainability points), reinforcing everyday rules aligned with policy objectives.

74 The second type of SG (Testing Retention) can strengthen the delivery component of programmatic success by verifying that trained procedures and decision rules are applied under pressure, constraints, and uncertainty, thereby producing traceable performance data (e.g., timing, sequence accuracy, error rates, coordination) that can reveal gaps before rollout and support targeted remediation. One example is "HazMat: Hotzone" (Williams-Bell et al., 2015), targeting frontline emergency responders who must execute protocols in chaotic scenes. The gameplay mirrors a hazardous-materials incident (identify substance; establish hot–warm–cold zones; select PPE; set up decontamination; manage evacuation) under time pressure, while the system captures step order, classification accuracy, throughput, and time-to-safe-scene. Another example is "EDGE Virtual Training" (Bink & Miller, 2015). Designed for interoperating services (law enforcement, fire, dispatch), it places multiple agencies in a shared incident to practice role-true coordination and communication; crucially, it enables

replay and after-action review from participants' perspectives to examine interagency handoffs and refine joint procedures.

75 **Crash-testing mechanism** SG likewise can increase the likelihood of programmatic success by providing information needed to (1) formulate interventions that are effective in terms of outcomes and (2) stress-test policy systems under shocks (disasters, attacks, demand surges), exposing vulnerabilities, bottlenecks, and blind spots.

76 For example, the "Murray–Darling Basin Game" supports water-governance design by combining stakeholder roles with hydrological modelling and production functions grounded in real datasets (Moore et al., 2011). Participants test allocation rules and instrument bundles against goals (river health, economic prosperity, metropolitan supply security), revealing emergent effects from coupled human–biophysical dynamics. Similarly, "SIMPORT-MV2" was created for infrastructure managers to support long-term decision-making on the Port of Rotterdam Maasvlakte 2 expansion (Mayer et al., 2005): players rehearse investment sequencing, commercialization, and environmental safeguards, observing how choices shape performance, side-effects, and risks.

77 Applied to existing arrangements, Crash-testing Mechanisms SG can assess whether systems, standard operating procedures, infrastructures, and inter-institutional handoffs hold up under extreme conditions without real-world risk. An example is "Operation Digital Ant", designed to diagnose food-supply chain security (Hofmeier & Lechner, 2021). Teams construct insider roles, attack chains, and countermeasures mapped onto realistic boards (e.g., meat-processing plant; logistics operator), producing stakeholder profiles, plausibility-rated attack scenarios, and analyses of countermeasures that can inform risk reviews, procedural updates, and new safeguards.

78 Finally, **Exploring Systems** SG can be useful across formulation, implementation, and evaluation by making system dynamics visible, surfacing context constraints, and engaging stakeholders in solution design. In doing so, they help align interventions with delivery constraints and implementation capacity. First, these games can help policymakers and implementers understand complexity (feedback loops, thresholds, side effects) before commitments are embedded in policy documents. For instance, the "Edible City Game" familiarizes urban policymakers and implementers with competing perspectives in urban food/green-infrastructure planning (Pueyo-Ros et al., 2023). Using a 3D model of Girona, participants adopt stakeholder roles with goals and authorities, negotiate, and observe impacts via dynamic indicators; the platform records action logs and indicator trajectories that can be exported into planning briefs or monitoring tables.

79 Second, this type of games can diagnose misfit between proposed measures and local practices and reveal behavioral tendencies that matter for implementation. An example comes from Senegal (d'Aquino & Bah, 2013): a companion-modelling approach combined an open board/role-playing game with an agent-based computer model to co-design and test land-use rules for climate-change adaptation. Participants (from ministries to land users) build scenarios, define indicators, play, jointly assess results, redesign rules, and test again, confronting top-down procedures with local constraints and iteratively adjusting sequences and rules toward an implementable framework.

80 Third, by involving stakeholders and target groups in evaluating or redesigning interventions, Exploring Systems games can surface needs and constraints of policy takers and support adaptation of measures. For instance, "The Food in Later Life Game" explores food services for older adults (Wills & Dickinson, 2023) using ethnography-based scenarios. Participants from health care, social services, NGOs, and business discuss barriers and operational frictions (e.g., transport hurdles, referral bottlenecks, staffing pressures) and jointly define practical standards (acceptable travel time, minimum information sets), producing concrete service-change shortlists that can inform modification of interventions.

SG and political success

81 The third dimension of policy success is political success i.e., the effect of a policy on incumbents, gauged by whether it strengthens their governing capacity, sustains the values they promote, and eases the "business of governing." Marsh and McConnell (2010) conceptualize this in relation to government and its agency (mandate, reputation, direction), rather than to "politics" in society at large. Correspondingly, three interrelated criteria apply: (1) reputation and electoral mandate, when the policy builds credibility and helps maintain or improve electoral prospects; (2) agenda control and governability, when the measure dampens conflict, reduces the costs of defiance, and allows attention to shift to other priorities; and (3) coherence of governmental course, when the action reinforces the overarching trajectory of values and program, or credibly signals a change in that trajectory. Assessments typically fall on a continuum, from strengthened agency to contested or precarious success.

82 Against this backdrop, the scope for SG to deliver political success is necessarily limited, since outcomes along this dimension are strongly shaped by factors beyond any single policy (macroeconomic conditions, geopolitics, exogenous shocks). Nevertheless, SG can operate as supportive instruments in specific niches where legitimacy, public trust, and acceptance can be reinforced, for example by structuring stakeholder dialogue and negotiation, by eliciting concerns and acceptance conditions in a safe deliberative setting, and by communicating policy rationales in interactive, experiential formats.

83 In this context, the potential use of **Exploring system** SG in achieving political success may be viewed as three-fold. First, they can serve as tools for engaging diverse stakeholder groups in simulated decision-making processes, thereby fostering shared understanding of policy objectives, facilitating dialogue and negotiation, and helping to resolve conflicts of interest. By enabling participants to explore different perspectives and interact constructively, such games can strengthen mutual trust, enhance the legitimacy of public policies, and increase their acceptance among the public. The "Multi-Hazard Tournament" may illustrate this function of SG. In the game, mixed teams from federal, state, and local agencies, NGOs, producers, and universities plan for a real basin across timed rounds, justify their choices (e.g., short "press releases"), and receive expert and peer feedback under shared budgetary and hazard constraints (Bathke et al., 2019). By requiring teams to articulate objectives, compare trade-offs, and revise plans publicly, the format can build a shared picture of impacts and strengthen perceived legitimacy and acceptance of candidate policies.

84 Second, SG can operate as a form of "field reconnaissance" to identify citizens' concerns regarding planned or already implemented policies. Games can elicit qualitative concerns, priorities, and acceptance thresholds that are harder to surface in conventional formats, helping decision-makers anticipate reactions and communicate or adjust measures. One of the games that demonstrate such potential is "Playing with Uncertainty" (Avendaño-Uribe et al., 2022). This game was designed for community-based resilience planning in Christchurch, New Zealand. Participants (community leaders, infrastructure operators, municipal staff) negotiate how to allocate limited budgets across social and infrastructure resilience while a shared dashboard reveals the consequences of their choices across rounds. Role-true decisions and facilitated debriefs surface concrete worries, priorities, and thresholds for acceptance, enabling local policymakers to anticipate reactions, adjust measures, and communicate more effectively, thereby supporting political legitimacy.

85 Third, SG can be used to inform citizens about government-led policies and promote them to the public. When widely accessible (for example, as a mobile application) they could complement or even replace traditional communication campaigns conducted through mass media, social media, or outdoor advertising. Their advantage over conventional methods lies in offering interactive, experiential engagement with policy content, enabling players to explore scenarios, make decisions, and observe the consequences of different actions. Such experiential engagement can strengthen understanding of policy rationales and, in some contexts, support trust and acceptance. An illustrative example of how a SG can promote a government policy is "PeaceMaker" video game (Gonzalez & Czlonka, 2010). This political simulation places the player in the role of either the Prime Minister of Israel or the President of the Palestinian Authority, with the objective of achieving a peaceful resolution to the Israeli–Palestinian conflict[3]. Players make strategic decisions in diplomacy, security, and economic policy, while real news footage and documented events immerse them in the political, social, and security realities of the conflict. By simulating the challenges, constraints, and trade-offs faced by political leaders, "PeaceMaker" offers citizens an opportunity to better understand the rationale behind government actions and the complexity of peace-building efforts. One of the objectives of the game was to foster the potential to enhance public trust in the political leadership's chosen strategies.

Practical implications

86 Discussion shows that the relationship between SG families and policy success is not uniform. The linkages are uneven: different families are defensible for different types of policy work and different dimensions of success. Table 5 therefore summarizes directional, practice-oriented alignments rather than an exhaustive mapping of "all games to all success criteria."

Table 5. A summary of the potential role of SG (by types) in contributions to policy success.

Policy success dimension	Criterion	SG type

Process	Legitimation and inclusiveness (representative engagement in consultation)	Exploring Systems
	Coalition-building (compromise around bundles of measures, sequencing, acceptance conditions)	Exploring Systems
	Procedural transparency (auditable deliberation trail)	Exploring Systems
Programmatic	Implementation fidelity (operational success)	Testing Retention
	Outcome attainment (effectiveness)	Teaching Knowables
	Outcome attainment and resource success (effectiveness and efficiency)	Crash-testing Mechanisms
	Outcome attainment & actor/interest success (effectiveness and distribution/acceptance)	Exploring Systems
Political	Reputation/mandate (trust & acceptance)	Exploring Systems
	Agenda control and governability (anticipate reactions, manage conflict)	Exploring Systems
	Coherence of governmental course (communicating rationale, aligning with values/program)	Exploring Systems

Source: own elaboration.

87 The analysis and illustrative cases discussed above point to three practical implications. First, selection should start from the policy task, not from the tool. Treating "SG adoption" as a single decision is misleading: Exploring Systems formats are most defensible when the primary need is to structure deliberation, make trade-offs explicit, and support process legitimacy; Teaching Knowables and Testing Retention are better suited to programmatic success when implementation depends on shared procedural understanding, handoffs, and repeated rule-guided behavior; Crash-testing Mechanisms are most appropriate when teams need early diagnosis of vulnerabilities, comparisons of instrument bundles, and calibration of sequencing and monitoring thresholds. In practice, this means specifying the decision need up front (deliberation vs. implementation support vs. design stress-testing) and choosing the family accordingly.

88 Second, policy relevance depends on what remains after play. Across the examples, games become policy instruments when they generate usable outputs that can be taken up in routine policy work. For process-oriented uses, these outputs are typically deliberative artifacts (e.g., clarified trade-offs, option bundles, acceptance conditions, and traceable rationales). For programmatic uses, they are operational artifacts (e.g., shared procedures, checklists, handoff protocols, performance traces, after-action review notes). For crash-testing uses, they are diagnostic artifacts (e.g., scenario libraries, vulnerability/failure-mode lists, candidate instrument bundles, and early-warning thresholds). Practically, this implies that game design should include an

explicit "output plan": define in advance which artifacts will be produced, how they will be documented, and where they will be used (e.g., consultation records, implementation roadmaps, monitoring plans, or updates to standard operating procedures).

89 Third, facilitation and transparency are safeguards, not add-ons. The cases discussed above imply that SG outcomes are easiest to defend when facilitation limits domination and keeps deliberation aligned with the policy question, and when assumptions, rules, and indicators are made explicit so that outputs remain interpretable beyond the session. This is particularly important because many game-derived insights are scenario-dependent: they support judgment and coordination but should not be treated as definitive evidence on their own. In practice, teams should budget for facilitation capacity, document assumptions, and decision rules, and triangulate game outputs with statutory consultation, legal due process, and evaluation routines before institutionalizing conclusions.

Linking games with policy reality

90 Transfer from game to policy reality is not automatic. There are three key aspects that determine the effective transfer between game-based learning and real-life applications: (1) validity of game design, (2) quality of debriefing, and (3) timing of the game exercise in the policy process.

91 The impact of a game on real-life scenarios is significantly enhanced when the game is aligned with reality. Literature identifies three dimensions of validity that contribute to this effectiveness (Peters et al., 1998):

- structural validity, which refers to how isomorphic the game's architecture is with the system it represents: that is, whether the actors, information flows, rules, and relationships in the game map meaningfully onto those in the real world;
- psychological validity, which speaks to players' own sense of realism: does the game feel authentic to them, and does it connect with how they actually behave outside of it?;
- process validity, which concerns behavior in action, whether the decisions, interactions, and choices players make during the game resemble those they would make (or do make) in the reference system.

92 Enhancing each type of validity may involve specific strategies, but overall, validity can be assessed and improved through repeated playtesting with intended users, along with transparent documentation of any underlying assumptions and limitations.

93 Debriefing in the context of serious games is a structured reflective discussion that takes place following gameplay. Its purpose is to extract tacit reasoning, cognitive strategies, and emotional responses from participants, ultimately fostering a connection between the in-game experiences and real-life practices. A well-designed debriefing process guides players from examining "what happened on the board" to understanding "what this tells us about the reality we want to change." Effective debriefings typically include structured reflection on strategies and pain points, collective reconstruction of observed system dynamics and feedback, explicit mapping of in-game experiences to relevant policies, institutions, and actors, and identification of actionable next steps. The game design literature provides numerous important

insights for structuring and facilitating debriefing sessions (Crookall, 2010; Kortmann & Peters, 2021; Tosterud et al., 2020).

94 Finally, the timing of a game exercise within the policy process is key for maximizing its relevance and impact. A game integrated at the right moment can effectively illuminate critical issues, facilitate consensus-building, and enhance decision-making. For instance, implementing a game during the early stages of policy formulation can provide stakeholders with a platform to explore different scenarios, test assumptions, and understand potential outcomes before committing to specific policies. Conversely, using a game during the implementation phase can serve as a tool for evaluating progress, addressing challenges, and adapting strategies as necessary.

95 Moreover, aligning the timing of the game with ongoing discussions, debates, or crises can significantly heighten engagement and investment from participants. Ultimately, a thoughtfully timed game exercise not only supports strategic policymaking but also fosters stakeholders' willingness to engage actively with the findings and implications derived from the gameplay experience.

96 To summarize, the interplay of game design validity, debriefing quality, and strategic timing collectively enhances the likelihood that insights from serious games will translate effectively into policy and management applications.

Conclusions

97 SG have become increasingly visible at the intersection of public policy, education, and governance practice. In response to our first research question on trends in SG use in public policy literature, our evidence shows clear growth of SG research across adjacent fields, yet a striking under-representation in core public policy journals. This looks less like a lack of mature applications and more like a diffusion gap between communities: game research develops and circulates largely outside the publication venues that shape mainstream policy debates. The practical implication is that SG already exist as a rich methodological repertoire, but policy scholarship still treats them as peripheral, which slows cumulative learning about when they help and when they mislead. As a result, the trajectory is uneven: SG research is expanding, but it is only weakly translated into policy-facing outlets where lessons could accumulate and travel across sectors and jurisdictions.

98 Against that background, the practitioner-facing evidence adds an important qualifier. Addressing our second research question on practitioners' preferences, rather than enthusiasm or rejection, we observe conditional interest: respondents tend to treat SG as potentially useful but not as a default instrument. Their reservations align with predictable constraints in public organizations such as time and capacity, accountability requirements, procurement and evaluation routines, but also with a more specific uncertainty about what, exactly, a game produces that can be defended and used after the session ends. Preferences are not uniform across game families, which suggests that "SG adoption" is not a single decision. Formats that help navigate interdependence and trade-offs appear easier to justify than formats that test retention and execution, where organizational cost, sensitivity, and evidentiary expectations are higher.

99 The central contribution of the article lies in linking a practice-oriented typology of SG, to the three-dimensional conception of policy success developed by McConnell (2010) and Marsh and McConnell (2010). This alignment clarifies why different SG families matter for different dimensions of success and what kind of contribution they can plausibly make through learning, coordination, sensemaking, and (in some cases) legitimacy-building, rather than as direct "drivers" of success. Teaching Knowables supports programmatic performance when implementation requires shared procedural understanding or when policy goals depend on repeated, rule-guided behavior change. Testing Retention speaks to implementation fidelity by revealing whether trained routines, handoffs, and decision sequences hold under pressure and uncertainty. Crash-testing Mechanisms helps teams probe failure modes early, compare instrument bundles, and calibrate sequencing and monitoring thresholds before interventions meet real-world complexity. Exploring Systems plays a distinctive role because it connects several dimensions at once: it can strengthen process success by structuring deliberation and making trade-offs explicit; it can support programmatic success by exposing context constraints and actor interactions that shape delivery; and it can offer bounded support for political success by helping decision-makers anticipate contestation and communicate rationales, while acknowledging that political outcomes are heavily shaped by forces beyond any single tool. Thus, regarding our third research question on the prospects of using SG for policy success, we may conclude that these prospects are conditional and format-specific: they are strongest when a game is designed around a clear decision need and when post-game artifacts can be integrated into routine policy work.

100 Across these distinct use cases, one condition consistently distinguishes SG that remain engaging events from SG that become policy-relevant instruments. Transfer from game to policy reality is enhanced by validity of game design, quality of debriefing, and timing of the game exercise in the policy process. SG can contribute through intermediate capacities (learning, coordination, sensemaking, and sometimes legitimacy), and only travel into practice when they are built around a concrete decision need and yield usable, interpretable outputs that persist beyond the session. The contribution of SG depends on whether play produces something that can be taken up in the institutional machinery of public action. Decision logs, scenario comparisons, risk registers, agreed-upon criteria, and concise rationales matter because they can enter impact assessments, implementation roadmaps, monitoring plans, and updates to standard operating procedures. Facilitation and transparency protect this transfer: skilled facilitation keeps deliberation aligned with the decision need and limits domination; clear assumptions and interpretable indicators protect credibility and reduce the risk of over-claiming. When these conditions are absent, SG may still help participants learn and coordinate, but the link to policy success becomes harder to sustain.

101 These claims also imply limits that should temper expectations. Many game-derived insights remain qualitative, structural, or scenario-dependent rather than statistically generalizable; poorly bounded models can create false confidence; and behavioral effects decay without reinforcement in real workflows. Moreover, the published evidence base remains skewed toward high-income settings, with comparatively few documented public-policy applications in Global South contexts. This limits claims of transferability and inclusiveness without careful adaptation. SG therefore complement,

rather than replace, statutory consultation, legal due process, and formal evaluation. Future research should address the transfer problem more directly but also broaden and systematize the evidence base in three concrete ways. First, survey and interview evidence should be replicated on a larger and more diverse sample of practitioners, allowing for comparisons across administrative tiers and organizational settings. Second, the field needs a structured, systematic review and inventory of SG used in public policy, paired with an assessment of their effects, especially their potential to mitigate policy failures by exposing assumptions, bottlenecks, and failure modes before they materialize. Third, SG should be analyzed comparatively across sectors and types of wicked problems to identify which game families travel best across domains and which are contingent on sector-specific constraints, contestation patterns, and accountability regimes. Building this agenda would allow cumulative learning not only about whether SG "work," but also about when they reduce the risk of failure, for whom, and through which institutional uptake mechanisms.

BIBLIOGRAPHY

Abt, C. (1969/1987). *Serious games*. University Press of America.

Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81*(2), 132–169.

Ansell, C., Sørensen, E., Torfing, J., & Trondal, J. (2024). *Robust governance in turbulent times*. Cambridge University Press.

Avendaño-Uribe, B., Lukosch, H., & Milke, M. (2022). Playing with uncertainty: Facilitating community-based resilience building. *Urban Planning, 7*(2), 278–294. https://doi.org/10.17645/up.v7i2.5098

Bathke, D. J., Haigh, T., Bernadt, T., Wall, N., Hill, H., & Carson, A. (2019). Using serious games to facilitate collaborative water management planning under climate extremes. *Journal of Contemporary Water Research & Education, 167*, 50–67. https://doi.org/10.1111/j.1936-704X.2019.03311.x

Bink, M. L., & Miller, J. T., II. (2015). *Assessing the use of game-based exercises in the Staff Attack-the-Network course* (Research Note No. 2015-02). U.S. Army Research Institute for the Behavioral and Social Sciences.

Boyle, A. E., Hainey, T., Connolly, M. T., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education, 94*(March), 178–192. https://doi.org/10.1016/j.compedu.2015.11.003

Cairney, P., & Geyer, R. (2017). A critical discussion of complexity theory: How does "complexity thinking" improve our understanding of politics and policymaking? *Complexity, Governance & Networks, 3*(2), 1–11.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical education, 42*(12), 1150–1152. https://doi.org/10.1111/j.1365-2923.2008.03172.x

Checkland, K., Hammond, J., Coleman, A., MacInnes, J., Mikelyte, R., Croke, S., Billings, J., Bailey, S., & Allen, P. (2023). "Success" in policy piloting: Process, programmes and politics. *Public Administration, 101*(2), 605–622. https://doi.org/10.1111/padm.12790

Çiftci, S. (2018). Trends of Serious Games Research from 2007 to 2017: A Bibliometric Analysis. *Journal of Education and Training Studies, 6*(2), 18–27. https://doi.org/10.11114/jets.v6i2.2840

Clark, J., Barton, B., Albarqouni, L., Byambasuren, O., Jowsey, T., Keogh, J., Liang, T., Moro, C., O'Neill, H., & Jones, M. (2025). Generative artificial intelligence use in evidence synthesis: A systematic review. *Research Synthesis Methods*, *16*(4), 601–619. https://doi.org/10.1017/rsm.2025.16

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*(2), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

Crookall, D. (2010). Serious Games, Debriefing, and Simulation/Gaming as a Discipline. *Simulation & Gaming, 41*(6), 898–920. https://doi.org/10.1177/1046878110390784

d'Aquino, P., & Bah, A. (2013). Land policies for climate change adaptation in West Africa: A multilevel companion modeling approach. *Simulation & Gaming, 44*(2–3), 391–408. https://doi.org/10.1177/1046878112452689

de Fine Licht, K. (2025). Resolving value conflicts in public AI governance: A procedural justice framework. *Government Information Quarterly, 42*(2), 102033.

de Graaf, G., Huberts, L., & Smulders, R. (2016). Coping with public value conflicts. *Administration & Society, 48*(9), 1101–1127. https://doi.org/10.1177/0095399714532273

De Paoli, S. (2024). Performing an inductive thematic analysis of semi-structured interviews with a large language model: An exploration and provocation on the limits of the approach. *Social Science Computer Review, 42*(4), 997–1019. https://doi.org/10.1177/08944393231220483

de Smedt, P., & Borch, K. (2022). Participatory policy design in system innovation. *Policy Design and Practice, 5,* 51–65. https://doi.org/10.1080/25741292.2021.1887592

Duffhues, J., Mayer, I. S., Nefs, M., & van der Vliet, M. (2014). Breaking barriers to transit-oriented development: Insights from the serious game SPRINTCITY. *Environment and Planning B: Planning and Design, 41*(5), 770–791. https://doi.org/10.1068/b39130

Duke, R. (2011). Origin and Evolution of Policy Simulation: A Personal Journey. *Simulation & Gaming, 42*(3), 342–358.

Duke, R., & Geurts, J. L. A. (2004). *Policy games for strategic management. Pathways into the unknown*. Dutch University Press.

Elias, G. S., Garfield, R., & Gutschera, K. R. (2012). *Characteristics of Games*. The MIT Press.

Flood, S., Cradock-Henry, N., Blackett, P., & Edwards, P. (2018). Adaptive and interactive climate futures: systematic review of 'serious games' for engagement and decision-making. *Environmental Research Letters, 13*(063005), 1–20. https://doi.org/10.1088/1748-9326/aac1c6

Fox, J., McKnight, J., Sun, Y., Maung, D., & Crawfis, R. (2020). Using a serious game to communicate risk and minimize psychological distance regarding environmental pollution. *Telematics and Informatics, 46*, Article 101320. https://doi.org/10.1016/j.tele.2019.101320

Fung, A. (2006). Varieties of participation in complex governance. *Public Administration Review, 66*(S1), 66–75.

Furber, A., Medema, W., & Adamowski, J. (2018). Assessing the Benefits of Serious Games to support Sustainable Decision-Making for Transboundary Watershed Governance. *Canadian Water Resources Journal, 43*(4), 401–415. https://doi.org/10.1080/07011784.2018.1477628

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441–467.

Gerring, J. (2007). *Case study research: Principles and practices.* Cambridge University Press.

Geurts, J. L. A., Duke, R. D., & Vermeulen, P. A. M. (2007). Policy gaming for strategy and change. *Long Range Planning, 40*(6), 535–558.

Gonzalez, C., & Czlonka, L. (2010). Games for peace: Empirical investigations with PeaceMaker. In *Serious game design and development: Technologies for training and learning* (pp. 134–149).

Goodspeed, R., Babbitt, C., Garcia Briones, A. L., Pfleiderer, E., Lizundia, C., & Seifert, C. M. (2020). Learning to manage common resources: Stakeholders playing a serious game see increased interdependence in groundwater basin management. *Water, 12*(7), 1966. https://doi.org/10.3390/w12071966

Hartley, K. (2019). Practicing public policy in an age of disruption. *Policy Design and Practice, 2*(2), 163–179.

Hassan, L. (2020). A review of the literature on gamification of e-participation. *Government Information Quarterly, 37*(3), 101461.

Head, B. W. (2019). Forty years of wicked problems literature: Forging closer links to policy studies. *Policy and Society, 38*(2), 180–197. https://doi.org/10.1080/14494035.2018.1488797

Hofmeier, M., & Lechner, U. (2021). Operation Digital Ant: A Serious Game Approach to Collect Insider Threat Scenarios and Raise Awareness. In Proceedings of the 2021 European Interdisciplinary Cybersecurity Conference (EICC '21). Association for Computing Machinery, New York, NY, USA, 14–19. https://doi.org/10.1145/3487405.3487655

Huitema, D., Jordan, A., Munaretto, S., & Hildén, M. (2018). Policy experimentation: Core concepts, political dynamics, governance and impacts. *Policy Sciences, 51*(2), 143–159. https://doi.org/10.1007/s11077-018-9321-9

Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. Prentice-Hall.

Kortmann, R., & Peters, V. (2021). Becoming the Unseen Helmsman - Game facilitator competencies for novice, experienced, and non-game facilitators. *Simulation & Gaming, 52*(3), 255–272. https://doi.org/10.1177/10468781211020792

Krpan, D., & Urbaník, M. (2024). From libertarian paternalism to liberalism: behavioural science and policy. *Behavioural Public Policy, 8*(2), 300–326. https://doi.org/10.1017/bpp.2021.40

Lanezki, M., Siemer, C., & Wehkamp, S. (2020). "Changing the Game—Neighbourhood": An energy transition board game, developed in a co-design process: A case study. *Sustainability*, *12*(24), 10509. https://doi.org/10.3390/su122410509

Marsh, D., & McConnell, A. (2010). Towards a framework for establishing policy success. *Public Administration*, *88*(2), 564–583. https://doi.org/10.1111/j.1467-9299.2009.01803.x

Maxim, R. I., & Arnedo-Moreno, J. (2025). Identifying Key Principles and Commonalities in Digital Serious Game Design Frameworks: Scoping Review. *JMIR Serious Games, 13*, e54075. https://doi.org/10.2196/54075

Mayer, I., Chin, R., Verbraeck, A., Bekebrede, G., van Schuijlenburg, M., Meijer, A. K., & Koeman, J. W. (2005). *Second Maasvlakte: Practice planning and design through simulation-gaming*. Delft University of Technology.

Mayer, I., Warmelink, H., & Zhou, Q. (2016). The Utility of Games for Society, Business, and Politics. A Frame-reflective Discourse Analysis. In N. Rushby & D. Surry (Eds.), *The Wiley Handbook of Learning Technology* (pp. 406–435). Wiley.

McConnell, A. (2010). *Understanding policy success: Rethinking public policy*. Palgrave Macmillan.

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

Memmert, L., Cvetkovic, I., Tavanapour, N., & Bittner, E. (2025). Brainstorming with a generative language model: Effect of exposure to AI ideas on brainstorming performance and cognitive load. *Business & Information Systems Engineering*. https://doi.org/10.1007/s12599-025-00974-y

Moore, G. A., Langford, J., Ayre, M., Learmonth, G., Brizga, S., & Wallis, P. J. (2011). The Murray-Darling Basin Game – A model to explore water allocation decisions. *19th International Congress on Modelling and Simulation*, Perth, Australia, 12–16 December 2011.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education, 15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

OECD. (2025). *Governing with Artificial Intelligence*. OECD Publishing.

Olejniczak, K., Wolański, M., & Widawski, I. (2018). Regulation crash-test: applying serious games to policy design. *Policy Design and Practice, 1*(2), 194–214. https://doi.org/10.1080/25741292.2018.1504372

Olejniczak, K., & Widawski, I. (2024). Serious games for evaluation. In K. E. Newcomer & S. Mumford (Eds.), *Research Handbook on Program Evaluation* (pp. 608–628). Edward Elgar Publishing. https://doi.org/10.4337/9781803928289.00041

Olejniczak, K., Borkowska-Waszak, S., Domaradzka-Widła, A., & Park, Y. (2020a). Policy labs: The next frontier of policy design and evaluation? *Policy & Politics*, 48(1), 89–110. https://doi.org/10.1332/030557319X15579230420108

Olejniczak, K., Newcomer, K., & Meijer, S. (2020b). Advancing Evaluation Practice With Serious Games. *American Journal of Evaluation*, *41*(3), 339–366. https://doi.org/10.1177/1098214020905897

Oyamada, S., & Kitani, S. (2019). Consensus building gaming which promotes creative solutions to wildlife management dilemmas. *Simulation & Gaming, 50*(5), 556–571. https://doi.org/10.1177/1046878119867366

Peters, V., Vissers, G., & Heijne, G. (1998). The Validity of Games. *Simulation & Gaming, 29*(1), 20–30. https://doi.org/10.1177/1046878198291003

Petko, D., Schmid, R., & Cantieni, A. (2020). Pacing in serious games: Exploring the effects of presentation speed on cognitive load, engagement and learning gains. *Simulation & Gaming, 51*(2), 258–279. https://doi.org/10.1177/1046878120902502

Pueyo-Ros, J., Comas, J., Säumel, I., Castellar, J. A. C., Popartan, L. A., Acuña, V., & Corominas, L. (2023). Design of a serious game for participatory planning of nature-based solutions: The

experience of the Edible City Game. *Nature-Based Solutions, 3*, 100059. https://doi.org/10.1016/j.nbsj.2023.100059

Reinhard, P., Li, M. M., Fina, M., & Leimeister, J. M. (2025). Fact or fiction? Exploring explanations to identify factual confabulations in RAG-based LLM systems. *CHI Conference on Human Factors in Computing Systems (CHI EA '25)*, Article 274, pp. 1–13. ACM. https://doi.org/10.1145/3706599.3720249

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155–169. https://doi.org/10.1007/BF01405730

Sawyer, B., & Rejeski, D. (2002). *Serious Games: Improving Public Policy Through Game-Based Learning and Simulation*. Woodrow Wilson International Center for Scholars.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, "translations" and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–1939. *Social Studies of Science, 19*(3), 387–420.

Szatkowska, W., Ćwil, M., & Wardaszko, M. (2025). Awareness over knowledge: An exploratory study on a refugee crisis game. *Simulation & Gaming, 56*(2), 210–236. https://doi.org/10.1177/10468781241311627

Tan, H., Yap, C. W., Heng, W. Y., Koh, E. H. J., Ho, W. Y., & Lim, Y. J. (2020). Healthene™: Developing a board game on value-based healthcare financing. *Simulation & Gaming, 51*(1), 87–105. https://doi.org/10.1177/1046878119888710

Taylor, L. (2021). Public actors without public values: Legitimacy, domination and the regulation of the technology sector. *Philosophy & Technology, 34*(4), 897–922.

Thomas, L. D. W., Romasanta, A. K. G., & Pujol Priego, L. (2026). Jagged competencies: Measuring the reliability of generative AI in academic research. *Journal of Business Research*, *203*, 115804. https://doi.org/10.1016/j.jbusres.2025.115804

Tosterud, R., Kjølberg, K., Kongshaug, A. V., & Haugom, J. V. (2020). Exploration of Two Different Structures for Debriefing in Simulation: The Influence of the Structure on the Facilitator Role. *Simulation & Gaming*, *51*(2), 243–257. https://doi.org/10.1177/1046878120903467

Tyler, T. R. (2006). *Why people obey the law* (2nd ed.). Princeton University Press.

Vecchio, L. P., & Del Greco, A. (2023). Game-based solutions and the plastic problem: A systematic review. *Sustainability, 15*(6), 5558. https://doi.org/10.3390/su15065558

Wilkinson, P. (2016). A Brief History of Serious Games. In R. Dörner, S. Göbel, M. Kickmeier-Rust, M. Masuch, & K. Zweig (Eds.), *Entertainment Computing and Serious Games* (pp. 17–41). Springer. https://doi.org/10.1007/978-3-319-46152-6_2

Williams-Bell, F. M., Kapralos, B., Hogue, A., Murphy, B. M., & Weckman, E. J. (2015). Using Serious Games and Virtual Simulation for Training in the Fire Service: A Review. *Fire Technology*, *51*(3), 553–584. https://doi.org/10.1007/s10694-014-0398-1

Wills, W., & Dickinson, A. (2023). Vulnerability to food insecurity among older people: The role of social capital. *Sociological Research Online, 28*(4), 946–963. https://doi.org/10.1177/13607804231170516

World Bank. (2020). *Global experiences from regulatory sandboxes*. World Bank.

NOTES

1. Search formula: TS=("serious gam*"), all languages, all document types, 1900–2025
2. This value, as well as those mentioned later in the text, refer to mean scores. Analysis of variance (ANOVA) indicates that the observed differences are statistically significant.
3. At the time of *PeaceMaker*'s release in 2007, the political context was significantly different from today.

ABSTRACTS

Public policy practitioners increasingly face complex, value-laden problems where the conditions for success or failure are difficult to anticipate and even harder to test in advance. This article examines serious games (SG) as a practical method for exploring these conditions before high-stakes commitments are made. Drawing on three literature reviews and a survey of 143 policy practitioners conducted across multiple institutional settings between 2020 and 2025, we ask: what is the current trend in SG use within public policy scholarship, what are practitioners' preferences regarding different game types, and what are the prospects for deploying SG in the context of policy success? Our findings reveal a striking diffusion gap: while SG research has grown substantially across adjacent disciplines, it remains largely absent from mainstream public policy journals, despite growing practitioner interest. Building on these findings, we apply an existing typology, which distinguishes four categories of SG: Teaching Knowables, Testing Retention, Crash-testing Mechanisms, and Exploring Systems, and we position each type in relation to the Marsh's and McConnell's process, programmatic, and political dimensions of policy success. We argue that SG offer a scalable, low-cost complement to existing experimentation infrastructures, particularly in settings marked by uncertainty, value conflict, and the ever-present risk of policy failure.

INDEX

Keywords: serious games, public policy, policy success, policy failure

AUTHORS

DOMINIKA WOJTOWICZ

https://idref.fr/268526540

Kozminski University, Department of Economics, Sustainable Development Center, Poland.
https://orcid.org/0000-0002-8433-8935

KAROL OLEJNICZAK

https://idref.fr/268526575

SWPS University, Center for Policy Design and Evaluation
https://orcid.org/0000-0002-7079-2440

TOMASZ KUPIEC

https://idref.fr/268526486

University of Warsaw, Centre for European Regional and Local
https://orcid.org/0000-0002-6469-7746