31. Serious games for evaluation

Karol Olejniczak and Igor Widawski

Communities worldwide face unprecedented challenges, whose breadth and urgency are well recognized in the Sustainable Development Goals (SDGs; https://sdgs.un.org/goals). In addressing these challenges, governments and societies use public policies to transform individual and collective behaviors and reshape systems (Dryzek & Pickering, 2019; Esty, 2019; Dixson-Decleve et al., 2022).

Evaluation craft, with its practical focus on determining "what works, for whom, and why," seems well-positioned to provide substantial assistance to policy designers and practitioners across the globe (Pawson, 2013; Patton, 2019). However, new challenges require an upgraded toolbox. In particular, evaluators need methods that allow questioning established development paradigms, safely experimenting with unusual policy options, anticipating unintended side effects, and building a shared understanding of wicked problems and their change mechanisms.

In this chapter, we demonstrate how serious games have the potential to serve evaluation in this transformative time. Serious games are games used for purposes other than entertainment (Abt, 1987; Flood et al., 2018). They have characteristics aligned with evaluation needs, cover a wide spectrum of policy topics, and dynamically develop with analog means and cutting-edge technologies. However, the evaluation literature rarely discusses the application of games. This is a knowledge gap and an opportunity worth exploring.

Our chapter aims to tap into this opportunity by offering an overview of perspectives on, and emerging practices for, using serious games for public policy and program evaluation. Our discussion is based on systematic literature reviews on game design theories and practices (conducted in 2017, 2019, and 2022), discussions of topical panels at international conferences of related professional associations (American Evaluation Association 2017, Australasian Evaluation Society 2018, International Simulation and Gaming Association 2019, International Conference on Public Policy 2021), and our own experiences as evaluation practitioners and game designers.

The remainder of this chapter is structured into four parts. First, we provide a brief overview of the origins of serious games, and we present the key characteristics of games that make them especially valuable for evaluative inquiries. The second part discusses a spectrum of possible applications of serious games in evaluation practice. We identify four major areas of applications, and we illustrate them with real-life examples. In the third part of the chapter, we dive into practicalities of the game development process for use in evaluation. We provide basic concepts and a framework useful for collaboration between evaluators and game designers, and we describe the typical steps of the design process. We conclude the chapter with guidance on using serious games in evaluation and suggestions for research and practice agendas. We hope this chapter will contribute to developing synergies and cooperation between the community of program evaluators and game designers worldwide.

SYNERGIES BETWEEN GAMES AND EVALUATION

The Origin and Definition of Serious Games

We can trace the origins of serious gaming from early forms of abstract war-themed games (e.g., chaturanga, chess, checkers), through seventeenth- and eighteenth-century military games, to advanced World War II operation simulations in modern times, followed by military and policymaking games in the 1950s (Duke, 2014; Wilkinson, 2016). The actual term "serious game" was introduced in 1969 by Clark Abt (1987) when describing games used for educational purposes for both adult and young learners. The practice was substantially advanced by Richard Duke, who promoted games as collective sensemaking that improves communication among competing stakeholders (Duke, 2011).

The use of serious games accelerated with an explosion of computer games in the 1980s. Most of those games focused on educational purposes, from digital puzzles for elementary schools, through vocational training and higher education, to collaborative workplace and training simulations (Djaouti et al., 2011; Dörner et al., 2016).

Contemporary serious gaming is a thriving practice. According to Metaari (a leading firm in the serious games industry), the sales of serious digital games will quadruple to $28.8 billion by 2025 (Metaari, 2021). Technological advances in games, particularly Virtual Reality (VR), drive growth. To this, we should add the commercial revival of analog board games, especially Eurogames (Woods, 2012). The creativity of board game designers, combined with new technologies like Augmented Reality (AR), create new possibilities for development.

Modern serious game applications are no longer limited to military crisis management or education. Games are used in health care, psychology, environment, engineering, business, and economics. The objectives of games also span well beyond teaching, to include behavioral change, safety testing, environmental and spatial planning, consensus-building, community integration, and social and civic change (for a systematic overview of applications, see Çiftci, 2018; Schrier, 2021; Stokes, 2020). In this chapter, we adopt a more elaborated definition of serious games that has been adapted for the field of public policy and evaluation (see Box 31.1).

BOX 31.1 DEFINITION OF SERIOUS GAMES

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.

Source: Olejniczak et al. (2020, p. 345).

A few aspects of this definition are worth noting. First, the definition covers both digital and analog games. We give special attention to analog games because they are cheaper in development, and they provide face-to-face interactions that are especially valuable in simulating policy decision-making processes. Second, what makes a serious game is the intent of use and

what happens after the game – that is, how the gameplay is translated into public policy. In practice, commercial games, initially designed for entertainment, can be adapted to specific contexts for evaluation goals. There is a wide space for innovation for games application, considering the size and variety of commercial game populations.

Third, the definition underlines the presence of clear rules, and distinguishes games from mere play. Fourth, by pointing out a use in a well-defined space, the definition distinguishes serious games from gamification initiatives. At the same time, this definition does not specify the level of abstraction used in games. Thus, it embraces both abstract games and simulations with high fidelity to reality. Finally, the definition points to the role of the game principal in deciding how the game could be used for real-life policy needs. In evaluation practice, this is likely a team of evaluators working with the public authority that commissioned the evaluation.

Usefulness of Serious Games for Evaluation

From the perspective of evaluation practice, certain characteristics of games make them valuable for evaluative inquiries. Games provide a laboratory for decision-making, a safe space for experimentation, a multilogue that engages stakeholders, a time machine for anticipating effects, and learning machines. Let us explain these five characteristics.

Typically, evaluations focus on improving decision-making in public policy and management. Games are a "series of interesting decisions" (Meier, 2020, p. 185). At their core, games are just like public policy: problem-solving with the use of limited resources. Like policy practitioners, players are problem solvers confronted with trade-offs (McGonigal, 2011). Moreover, when making a decision under time pressure with limited information, the players experience cognitive overload, and follow a sequence of choices similar to the theory of disproportionate information-processing in government (Baumgartner & Jones, 2005). Thus, we can see games as a microcosm for studying decision-making and collective problem-solving – a laboratory of decisions.[1]

When working for public institutions, evaluators must cope with the uncertainty that springs from the complexity of policy issues (Bamberger et al., 2015). Furthermore, the public sector is highly risk-averse, sometimes blocking evaluators' attempts to improve programs. Games could be used to mitigate those factors. Board games have often been used to investigate the theory of emergence, showing gradually "how complex behavior emerges from simple components" (Gobet et al., 2004, p. 6). Also, uncertain outcomes are at the core of a game (Stokes, 2020); in fact, exploring uncertainty in a non-threatening way is what makes a game fun (Costikyan, 2013). In that respect, games offer a so-called "magic circle" – a safe space for exploring wicked problems and experimenting with bold solutions (Huizinga, 1949; Klabbers, 2009).

In public policy, understanding different stakeholders' perspectives is essential for an inclusive, democratic process (Peters, 2018). Evaluation practice also strongly emphasizes inclusion, equity, and cultural responsiveness (see the American Evaluation Association's Guiding Principles: https://www.eval.org/About/Guiding-Principles; and Chapters 5, 6, and 9 in this *Handbook*). In this respect, games can be useful in two ways. First, they provide us with a communication language that is universal and non-threatening, engaging players in a "multilogue" mode of communication (i.e., a dialogue of multiple stakeholders aimed at consensus; Duke, 2011, p. 347). Furthermore, games provide an opportunity for participants in the policy process to change places and perspectives, and see the world through the eyes of other actors,

perhaps allowing them to overcome power imbalances and biases, build empathy, and share understanding (Edwards, 2023).

Evaluators are often challenged to anticipate the effects of proposed solutions. In complex reality, the effects of interventions usually occur with a long delay and/or are spatially misplaced. The time factor in a game can be manipulated to allow participants to observe feedback loops and side effects, consider longer-term consequences of their decisions, and reflect upon the sustainability of their design choices. Thus, game sessions can work as "time machines," helping evaluators and their program stakeholders see future effects.

Finally, evaluators are often challenged with limited uptake of evaluation findings and a mismatch between the demand for information and supply of study findings (Palenberg & Paulson, 2020; Weiss, 1988). Research indicates that games can effectively transfer different types of knowledge (like conceptual or procedural knowledge), skills (e.g., navigating social interactions, understanding values, managing emotions), and even capacities in the psychomotor domain (Kapp, 2012; Zhonggen, 2019). Games are well-aligned with human learning mechanisms, offering an experience that is emotionally engaging, social, experiential, and accompanied by immediate feedback (Boyle et al., 2016). In addition, good games often create among players a state of "flow," capturing their attention entirely (Csikszentmihalyi, 2008). This makes games "learning machines" (Beavis, 2017) – a useful tool for transferring knowledge and involving potential evaluation users.

AREAS OF APPLICATIONS

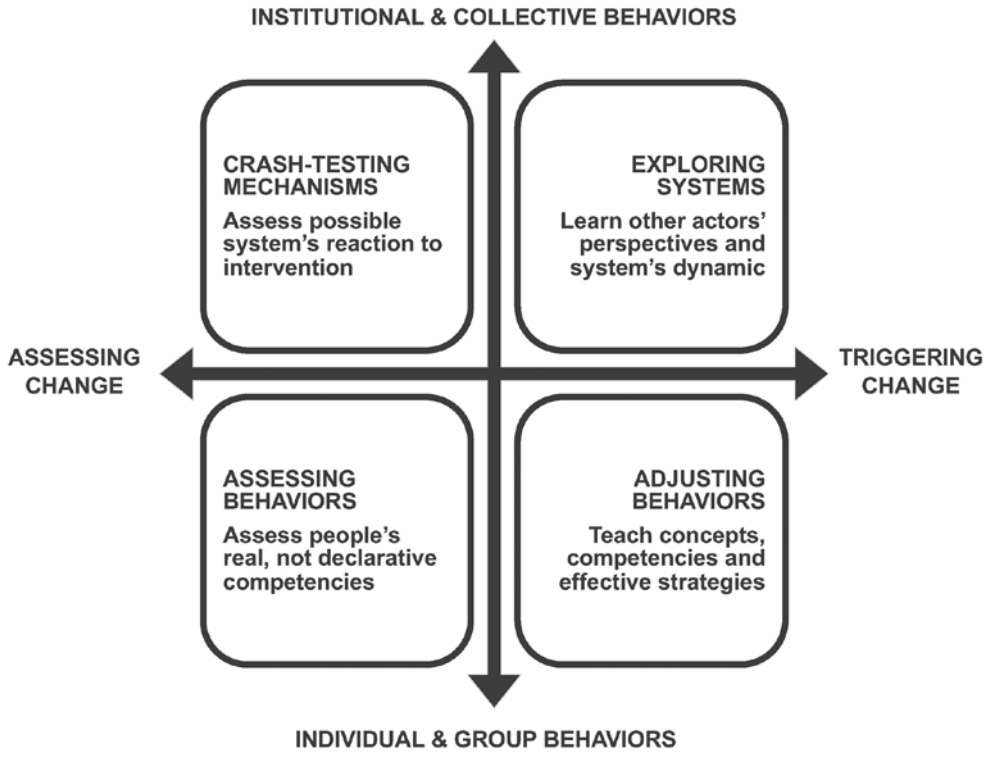
Overview of Areas of Applications

In order to organize the rich body of serious games and their potential spectrum of uses, we adapt the typology developed by Olejniczak et al. (2020). The potential areas of applications of games are organized around two major dimensions crucial to evaluation (see Figure 31.1): the evaluation purpose, and the nature of the evaluation subject (or evaluand).

The horizontal dimension in Figure 31.1 is aligned with the classic dichotomy between summative and formative evaluation purposes (Donaldson et al., 2010; Mathison, 2005; Scriven, 1991). The summative purpose is focused on *assessing change* among the players and systems they represent. The game is a method to measure a change caused by a program. These applications are connected with a more Positivist approach to evaluative inquiry, where the evaluator controls and manipulates the experimental factors, and observes the reactions of the subjects (players, in this case) (Rossi et al., 1999; Shadish et al., 2001; see also Chapter 20 in this *Handbook*).

When there is a formative purpose, the game is an intervention itself: it is focused on *triggering change* among the participants and the systems in which they are nested. These applications are aligned with a participatory, action research approach, where evaluators and policy stakeholders are engaged in interactions to co-create social change (Cousins & Whitmore, 1998; Patton, 2010; see also Chapter 7 in this *Handbook*).

The vertical dimension covers the object of the evaluative inquiry, the so-called "evaluand." Here we follow a perspective from both classic (Lasswell, 1951) and recent public policy literature (Peters, 2018; Shafir, 2013). It states that public policies are, at their core, about human choices and behaviors. For our matrix, we distinguish between *individual and group behav-*



Source: Adapted from Olejniczak et al. (2020, p. 346).

Figure 31.1 Areas of games application in evaluation

iors versus *institutional and collective behaviors*. The former are behaviors of individuals operating in small groups, and include games and evaluations dealing with skills, knowledge, choices, routines, and even psychomotor behaviors. The latter focuses on complex dynamics among organizations and collective actions, and covers situations when many policy actors interact with each other, addressing the infrastructural, environmental, and socio-economic factors that evolve.

Using the two dimensions together, we arrive at four broad areas of game applications in evaluation. These are (1) assessing behaviors, (2) adjusting behaviors, (3) crash-testing mechanisms, and (4) exploring systems (Table 31.1). In the next section, we illustrate these areas with specific examples of games. However, we need to emphasize that individual games are highly versatile. As we will see in the examples, minor adaptations in game session or content can provide for different game applications.

Examples of Four Areas of Applications

Area 1: Assessing behaviors

In the first area, the game is a method used to measure individuals' behaviors, and assess their skills, knowledge, competencies, or reactions to certain events. Players are put in certain roles and exposed to specific situations. Their responses are subject to the observation of evaluators, and behaviors in a game are assessed in line with preset benchmarks for specific skills and competencies. Evaluators can use such games to measure how a specific program under investigation is changing the capabilities of the target group. Evaluators can also use a game to collect data about the program target group, such as their initial level of competencies that are to be addressed by a program.

What is especially valuable for evaluation, is that this type of game works as a quasi-experiment: the behavior change is measured in a dynamic but still controlled environment. Thus, the behavioral responses are close to reality. Furthermore, evaluators can arrange separate game sessions with the program target group and a control group to have comparison points. When using this type of game, evaluators must pay attention to the construct validity of the game measures (i.e., to what extent does the game measure the phenomenon in question).

A typical example of assessment games is a PBS KIDS website containing a series of math games for children (https://pbskids.org/), accompanied by a learning analytics platform (LAP) (https://cresst.org/education/pbs-kids-big-data-for-little-kids/). The platform tracks the number of indicators throughout the gameplay of the children (e.g., correct responses, time spent on the task, etc.). Research has confirmed that those indicators significantly correlate with scores from a standardized math test (Roberts et al., 2016; Shute & Sun, 2020). Thus, evaluators could use games like these instead of tests (which are more stressful for the target group) to check the impact of new education programs on children's skills.

Another example is ProRail, the game executed for the Dutch railway administration to establish if traffic controllers can handle high-stakes scheduling changes. The game design accommodated both high- and low-tech solutions. On the one hand, it was a computerized simulation of train traffic flow, giving players control interfaces similar to the human–machine interface of a real train traffic control workstation. On the other hand, it was set in a typical office environment with real office dynamics in which control operators function (Meijer, 2015). The task for the players was straightforward: to open the bridge for the transport flow at optimal intervals. Assessing the capacities of individual staff was part of a larger policy decision-making process. The results of the game showed that experienced operators could handle extensive traffic. Furthermore, findings helped evaluators recommend the choice of policy solution: invest in training operators instead of building elevations for railway tracks.

The last example in this section is the PlaySmart video game for 16- to 19-year-old adolescents in a choose-your-own-adventure style (https://www.play2prevent.org/our-games/playsmart/). Players are characters placed into various situations, such as a party, and they are shown what happens if they make certain choices, such as kissing someone with herpes. The player is then taken through the consequences of that decision. This game uses a "time machine" option, so players are allowed to "go back in time" and make different choices (Smith, 2021). The game was developed by Lynn Fiellin from the Yale School of Medicine. Initially, it was applied to collect data on adolescents' perception of risk, particularly the harm of opioids. The data from the player's game decisions were used to better inform drug awareness and prevention programs.

Over time, the researchers observed that the PlaySmart game imparts players with information that can help them avoid drug misuse, and even promotes mental health and wellbeing more broadly. That modification transformed PlaySmart from "assessing behaviors" to our second application area, "adjusting behavior." The effectiveness of the PlaySmart game as an intervention is currently a subject of randomized controlled trials (Smith, 2021).

Area 2: Adjusting behaviors

Games in the second group focus on triggering positive change among individuals. Thus, the game is an intervention tool administered by game principals to teach individual players new skills, knowledge, and understanding of certain policy phenomena, and/or strategies for improving their behavior. These games can also be used for team capacity-building in organizations. In *ex ante* evaluation, evaluators can use a game to educate stakeholders about certain aspects important to the program topic, or introduce certain concepts and frameworks that could be used for discussing the program's theory of change. In an action-oriented approach, evaluators as co-implementers of the program can use the game as an intervention tool to improve the competencies and behaviors of the program target group. Finally, a game can be used as a mode of sharing evaluation findings.

The condition for successful use requires evaluators to have a well-structured understanding of the knowledge to be transferred by the game. The game design must effectively create engaging, problem-based, experiential learning for participants. In order to increase knowledge uptake, game sessions usually consist of game rounds interwoven with facilitated debriefings to help players reflect on their experience and link it with reality.

A simple example is the EAST card game (https://www.bi.team/publications/east-four-simple-ways-to-apply-behavioural-insights/). The British Behavioral Insights Team developed this game to inform busy policymakers about the key principles of applying behavioral insights to improve public programs. The game consists of a deck of cards with 32 specific behavioral techniques that have been shown to be effective in encouraging behaviors (e.g., checklists, chunking, salience, prompts, messenger effect) grouped into one of four colors representing: Easy, Attractive, Social, and Timely (Service et al., 2014). One player, a policymaker, proposes a policy challenge (e.g., reduce household food waste) and shuffles cards among other players. The players are to propose as many intervention ideas as possible using cards in their hand. At the end of the round, they present ideas, and the policymaker awards points. Then, the next player takes the role of policymaker. In the program evaluation context, teams could use this game to improve existing programs with behavioral insights, or brainstorm about new interventions.

In recent years, societies worldwide have been hit hard by two pandemics: the COVID-19 virus, and a virus of lies, fake news, and misinformation ("disinfodemic") (Awan et al., 2022; UN News, 2020). The "Bad news game" (https://www.getbadnews.com/en) was developed to "vaccinate" citizens against the virus of fake news. It does this in a slightly perfidious way. Players in a game take on the role of a fake news producer, mastering six techniques typically used in producing misinformation: polarization, invoking emotions, spreading conspiracy theories, trolling people online, deflecting blame, and impersonating fake accounts. Learning about those techniques in the game makes people better at spotting and resisting misinformation in real life (Roozenbeek & van der Linden, 2019).

There are also examples of games targeting experts' skills or attitudes on the topic of data and evidence use. Morewedge et al. (2015) reported that two computer detective games were

applied to reduce typical decision biases of intelligence analysts. Those games effectively educated analysts and evaluators about their own biases (e.g., spot bias, confirmation bias, fundamental attribution error, anchoring, representativeness, social projection; Barton et al., 2016).

Area 3: Crash-testing mechanisms

The primary purpose of the third set of games is to test, in close to real conditions, the reaction of institutional and collective actors to a policy solution (e.g., program, regulation) or event (e.g., emergency crisis). Players in these games are individuals or teams acting as system actors (e.g., representing specific institutions, organizations, or groups of stakeholders). During the game, players are exposed to new configurations of rules or a specific policy event; game principals manipulate system conditions, and then observe players' decisions and emerging strategies.

Program evaluators can use games like this to challenge pre-existing beliefs about what will work, particularly assumptions underlying the theory of change of specific complex programs or regulations. After running these games, evaluators understand possible mechanisms triggered among policy actors, and overall system response to changes. Running this type of game requires evaluators to design a game that includes the rules of play, positions, and resources assigned to the stakeholders of the intervention, as well as key contextual factors that could influence possible outcomes.

The Rubber Windmill is a classic example of this type of game application. The British Office for Public Management used the game to test a planned regulatory reform, and safely explore the unintended consequences of restructuring the UK's National Healthcare System. The players were representatives of different levels of government and healthcare systems, including hospitals. When confronted with a new financing logic, players revealed several shortcomings of the new policy design. The results of the game have been discussed in the British Parliament (Duke & Geurts, 2004).

Another example is the Kościerzyna rural transportation game (Olejniczak et al., 2018). It was created to evaluate, *ex ante*, a new law drafted by the Polish government in response to the decline of the Polish rural bus transport market and the decreasing accessibility to buses for residents in rural areas. During two days of game sessions, representatives of local municipalities, and big and small transportation companies, played their jobs on the region's fictional (but close to real) map, equipped with resources and competencies closely matched to the Polish reality. At a certain time in the game, players were confronted with new rules governing competition on public service contracts. Players were tested in the game about tender procedures that would take months to be deployed in real life, and would not be visible for at least a year due to the annual budgetary cycle. Players' responses, adaptation, the rationality behind their strategies, and changes in the model region map, provided data for evaluators to assess the likely effectiveness of the proposed law. The game also revealed that the assumption the new regulation would negatively impact smaller bus companies was groundless.

Games focused on the crash-testing mechanism also include a large number of simulations and decision games that deal with the emergency response of public organizations and local actors to natural disasters (McCreight, 2019), terrorist attacks, or health emergencies like pandemic (McGrady & Curry, 2021).

Area 4: Exploring systems

The final area of games application also deals with complex institutional and collective behaviors and, at first glance, looks similar to crash-testing mechanisms. However, there are two major differences. First, the challenges players face during the game are vaguer and broader, and tend to have unclear outcomes. Outcomes of decisions within a game are debated by players, and then announced by the game principal, or they emerge later, after the game. Second, the role of the players is different. When confronted with a challenge, players engage in conversations in which they focus on making sense of the policy issue while exploring each other's assumptions, complex relations (including biases and power imbalances), and eventually collectively develop a solution. Evaluators are facilitators of this "multilogue," while players are co-producers of collaborative problem-solving. Desired outcomes can include new strategies to deal with policy problems, and increased empathy and social capital among participating stakeholders. To summarize, games in this category simulate collective sensemaking about wicked problems.

As noted above, games designed for one purpose can be modified to accomplish additional objectives. For example, a game that simulates responses to a natural disaster is a crash-testing mechanism game, but it can be extended to simulate the recovery phase. In that case, players can begin more open-ended discussions and planning of programs and actions necessary for community rebuilding after the disaster, which turns the game into an exploring system type.

Similarly, the Kościerzyna transportation game was modified from being a crash-testing mechanism game into system exploration by simply adding a debriefing that involved players in a discussion of "what if" after the game. Participants compared and commented on each other's assumptions within the game, vested interests, and trust, and brainstormed changes in the tested regulation to improve its effectiveness. These add-ons transformed players from observed subjects to co-creators of policy solutions.

Recently, several game applications in this area have addressed climate adaptation decisions (i.e., adaptive governance). The case of the Marae-opoly game in New Zealand illustrates how researchers used the game to develop a respectful partnership with the Māori community facing difficult adaptation to climate change flooding hazards (https://www.fairplanet.org/story/can-a-board-game-support-climate-adaptation/). The game content was generated with hybridizing knowledge systems, combining traditional scientific data and detailed local knowledge. The game allowed the Indigenous community to negotiate new environmental and social realities, and develop their own place- and culture-appropriate adaptation options. The result was "mutually agreeable dynamic adaptive policy pathways (DAPP) for localized flood adaptation" (Blackett et al., 2022).

System exploration games can spark change in perceptions about outcomes, and induce change in local social networks, bridging divided communities, and strengthening a sense of place and group identity (see Stokes, 2020, for recent examples).

PRACTICALITIES OF THE GAME DESIGN PROCESS

Game design is a multi-step effort, requiring a combination of skills and high level of specialization. Designing useful board games requires the cooperation of game designers and graphic art experts. Digital game design is a true team effort, bringing coders, platform developers,

Table 31.1 Key characteristics of areas of games application in evaluation

	1. Assessing behaviors	2. Adjusting behaviors	3. Crash-testing mechanisms	4. Exploring systems
Game purpose and function	Assess degree of change in individuals or teams Game is a policy research method	Cause change among individuals or teams Game is a policy intervention tool	Assess change mechanisms among system's actors Game is a policy research method	Co-create system change with stakeholders Game is a policy intervention tool
Scope/focus of game	Individual skills, knowledge, behaviors	Individual skills, knowledge, behaviors	Institutional and collective behaviors of systems	Institutional and collective behaviors of systems
Relation between players and game principals	Players are subjects of observation conducted by game principals	Players are targets of intervention delivered by game principals	Players are subjects of observation during system manipulation led by game principals	Players and game principals are co-creators of change mechanisms
Expected output of the game sessions	Evaluators learn about the degree of individuals' capacities related to the policy issue	Individuals exposed to game gain skills and knowledge, and/or improve their behavior	Evaluators gain understanding about how actors in the system would respond to an intervention or event	System actors gain shared understanding of the system, develop empathy, and/or induce change in the system

Source: Authors' own work.

and even music composers to the process. Thus, serious game design calls for a partnership between evaluators and teams of game designers.

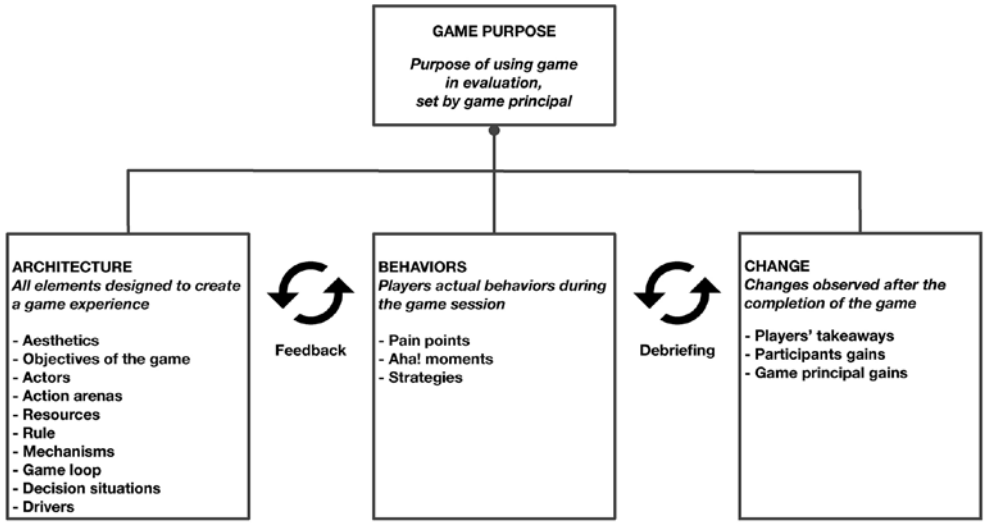
We recommend that evaluators interested in using games cooperate with game designers, rather than trying to develop their own expertise in games. However, productive cooperation requires a shared understanding of concepts and activities. In the following sections, we present main concepts useful in communication with game designers, and we discuss the process of game development.

Proposed Framework for Developing Serious Games

The game literature provides several guides on developing digital and board games (Adams, 2009; Brathwaite & Schreiber, 2008; Fullerton, 2019), and even frameworks to conceptualize game design (e.g., MDA framework developed by LeBlanc, 2004; system approach developed by Sellers, 2017). However, similar terms are sometimes used inconsistently, and game effects may be defined too narrowly, focusing on emotional impact and omitting changes in knowledge or behaviors.

Therefore, we propose a framework to help evaluators and game designers communicate during game development. In particular, we want to help evaluators articulate key choices in game design that could determine the future utility of specific games for evaluation. The framework uses some game design terminology, but links it with the public policy literature and adjusts to evaluation purposes discussed in the previous section. It can be applied when adapting existing commercial games, or preparing a blueprint for designing original games (both analog and digital).

The ABC framework is presented in Figure 31.2. Below we briefly discuss its overall logic. Appendix Table 31A.1 provides detailed definitions of all the terms. We envision it as a quick reference list for improved communication between evaluators and designers during game development.



Source: Authors' own work.

Figure 31.2 ABC framework for applying serious games to public policy purposes

The serious game is shaped by its purpose (see the previous discussion on four areas of games use for evaluation). Game architecture is composed of all elements that make a physical or digital game. Behaviors are actual, observable responses of players during their interactions within a given game architecture. And a change is a difference triggered by the game, and observable after participants complete the game.

The framework includes two mediating elements: feedback and debriefing. Feedback is a response to a trigger for players to reflect and adapt behavior during the game. In contrast, debriefing transfers the game experience beyond the game, linking observations from the game with reality. It is important to point out that feedback and debriefing may trigger changes. For example, points given to players in feedback come from existing rules, but a change in players' behaviors can also trigger new rules and options for points in the game. Also, debriefing by an evaluator can help link strategies in a game with real-life situations, and making this link can encourage players to try strategies inspired by reality within a game.

The ABC framework has a logic familiar to evaluators: it is structured as a Theory of Change. This includes the architecture of intervention (i.e., game), the behaviors that are supposed to be triggered among the target group, and the observable change desired.

Process of Game Development

Designing serious games, or adapting existing games to policy purposes, is highly iterative, with several loops focused on confronting ideas in a team and with future users. Designers jump between exploratory research and the action of design, and between details of reality and abstraction (Olejniczak et al., 2018).

To simplify, we can identify four phases of game design. We will discuss them briefly, showing the major decisions to be made, and illustrating with an example of a real game to

show how the choices unfold in practice. The example we use here is one of a game we developed called Knowledge Brokers (see Box 31.2).

BOX 31.2 KNOWLEDGE BROKERS BOARD GAME

The purpose of the Knowledge Brokers (KB) board game is to teach public policy analysts how to use the results of evaluation studies to effectively inform decision-makers. A one-day session is designed for up to 30 policy professionals working in teams of five to six. Players assume the role of knowledge brokers, with each team representing a fictional evaluation unit. Each team's objective is to provide decision-makers with expertise to help implement various socioeconomic projects. During every round, teams receive knowledge needs – that is, research questions concerning implemented projects. The correct response to the need requires a sequence of decisions: choosing an appropriate research design, identifying the key user of the study, and picking the best-tailored methods for delivering knowledge to users. The game is played under time and resource constraints. The winning team is the one with the highest record of utilized reports, which depends on the accuracy of their choices.

Knowledge Brokers was developed by two Polish companies: EGO (Evaluation for Government Organizations) and PGS (Pracownia Gier Szkoleniowych). The game has been used for training individuals – evaluators, applied researchers, staff commissioning evaluation, students of program evaluation – and for capacity-building with evaluation units in governments and international organizations.

Source: Olejniczak (2017).

Phase 1: Establishing the design brief

The goal of the starting phase is to establish the baseline conditions that would shape a blueprint of the game. Key decisions cover the purpose of the game for evaluation, initial preferences about game aesthetics (e.g., the form of the game – digital or analog; the narrative – real life or metaphor), and initial ideas about potential gains from the game session for the game principal (which is to a large extent aligned with the game's purpose). Game designers at this stage usually ask about such practicalities as the desired level of abstraction in a game (i.e., are we looking for a big picture or detailed simulation of reality), the length of the game, and the number of players that one game session might accommodate. All those ideas are summed up in a one-page design brief.

In the case of the Knowledge Brokers game, the purpose was clear: to educate players about factors and strategies that influence effective evaluation use. The players' takeaways were specific learning outcomes: (1) identifying knowledge needs, (2) acquiring credible knowledge with valid and relevant research designs, (3) applying effective methods to inform decision-makers, (4) strengthening policy arguments by combining findings from single studies into one knowledge stream, and (5) being aware that evaluation findings are only one factor in the decision-making process.

The rationale for selecting the analog form was twofold: it is cheaper and quicker in execution, and we wanted to promote in-person interactions and use the game sessions for networking among civil servants who usually work in isolation. At this stage, designers aligned aesthetics

with the topic of the game. Decisions in government are informed by pieces of evidence; in our game, players were building their reports like puzzles, with pieces of magnetic elements representing elements of a good report (including research questions, type of research design, identification of key users of the study, and methods for delivering knowledge).

Phase 2: Exploring and synthesizing the system

Phase 2 entails mapping the system to be recreated in the game. Game designers ask experts in the specific policy issue about key actors in the system, action arenas in which the policy takes place, resources available, and trade-offs. The main dilemma is to decide what should be reflected in the game, and what should be omitted. Discussion can be tense and challenging, but highly enlightening; evaluators may focus on specific details and push for complexity. At the same time, game designers will try to distill key elements of the system, so that future players can understand it sufficiently.

In Knowledge Brokers, after an initial discussion about recreating the whole system of monitoring and evaluation for European Union (EU) structural funds, we set the boundaries to focus only on evaluation units, and only on their knowledge production and dissemination work. What became an action arena (i.e., a board for the game) was a calendar with a lifecycle of projects to be aided by evaluations. The main constraints of our players were resources of time and staff required to supervise evaluation designs. With this narrowed focus, we developed a list of four key mechanisms that our players had to address with their decisions: (i) validity (matching research designs to questions), (ii) timing (providing answers on specific decision moments), (iii) relevance (matching knowledge with the needs of specific users), and (iv) accessibility (matching communication styles with types of users).

Phase 3: Designing architecture and feedback

This stage is devoted to meticulously constructing and combining game architecture elements into a working game. Usually, designers first define a sequence of major decision situations. Second, they develop a game loop. And third, they build around it all other elements of the architecture, adding or removing drivers to regulate game dynamics and fun factors.

Our understanding of the system of knowledge brokering was based on a systematic literature review and earlier studies on evaluation use (Olejniczak et al., 2016). The game loop followed typical decisions required for the evaluation execution (i.e., decide if you are addressing a knowledge need, choose an appropriate research design, identify the main user of the study, and pick methods for delivering knowledge). The topic of knowledge production seems quite static and uneventful. Thus, we introduced a whole package of drivers: achievements (completing 19 knowledge needs and delivering them to a jury), competition between teams (comparing scores each round), cooperation within teams (a division of work), countdown (ticking clock in every round), resource management (both in terms of labor and time), and even unexpected events (like health leave of personnel, an unexpected question from the minister, etc.).

Feedback was provided to players after each turn, but only to the groups that had completed their reports within the given deadline. Detailed feedback was given on a card with infographics and percentage points on how well the team satisfied four criteria: validity, timing, relevance, and accessibility for specific reports. Workshop facilitators use automated algorithms to assess the results quickly. The game algorithm includes a spectrum of good enough and optimal choices.

Phase 4: Testing the prototype

The goal of this phase is to test if the game is balanced, which entails running a series of pilot sessions (with paper prototypes) and observing players' behaviors in the game. This is followed by an honest, often painfully critical discussion with testers. Finally, developers may modify the game architecture and feedback mechanisms based on this feedback.

In the case of the Knowledge Brokers game, testing resulted in substantial changes to the game content. It turned out that a realistic toolbox of research and communication methods available for players was too complex, and decisions on choosing them were perceived to be too boring, long, and hard to follow. Thus, we reduced over 40 available research and communication methods to two smaller sets: eight research designs, and ten methods of providing knowledge to users. Another change was the wording of instructions and cases used in the game. In the dynamic gameplay situation, wordy texts were producing cognitive overload among players. Thus, expert jargon had to be removed, and the text had to be shortened.

Phase 5: Calibrating the game

In this phase, the game is finished and deployed to players. However, each game session can be used to improve the game. The goal is to tweak the game, based on data collected during sessions, to increase the game's effectiveness in fulfilling its purpose. At this stage, evaluators and designers usually continue improving debriefing, and pay special attention to capturing the changes generated by the game.

In the Knowledge Brokers game, after the first few game sessions, we changed the form and timing of debriefing. Instead of one debriefing at the end of the game, we provided three mini-lectures delivered every few rounds. Delivery of chunked information early in a game addressed the pain points experienced by players, and facilitated "aha moments" among players. Furthermore, it increased motivation throughout the game: players could immediately implement lessons from each mini-lecture to adapt their strategies and improve scores in the subsequent rounds.

The above discussion and our example of the Knowledge Brokers game illustrate the design process of a new game. The development team engaged all segments of the ABC framework. However, as we noted, evaluators can adapt existing commercial games. In that case, the game's architecture is already given. Evaluators can shape behaviors and change elements of the framework by modifying certain rules in the architecture of the original game, manipulating feedback, and introducing additional debriefings.

CHALLENGES IN GAME DESIGN AND APPLICATION

Based on our experiences and recent literature (Edwards, 2023), we identify five major challenges for game design and application in program evaluation. First, the design process is highly non-linear and can be time-consuming. Especially when designing new digital games, playtesting and bug detection can be cumbersome. Boardgame development takes two to six months, while designing digital games can take four to 12 months.

Second, balancing reality and abstraction in a game can be challenging. We want players to quickly see the big picture of the system and recognize their roles in the game, but we do not want them to fixate on all the missing details from the complex reality.

The third challenge is to secure the involvement of policy actors. Busy professionals are difficult to recruit for lengthy sessions. Also, people often do not see games as constituting a serious research activity. We usually try to address this by intriguing potential participants with the method (using words such as simulation rather than game), emphasizing insights they can gain, accepting smaller groups (allowing more interaction and opportunities to network), partnering with academic and public institutions to organize sessions, and aiming for shorter sessions (generally, two hours to half-day).

The fourth challenge is to ensure skillful facilitation and debriefing. Players are put in an unfamiliar situation; they can easily "get lost." Furthermore, although games are non-threatening, they can spark many emotions, competitions, and conflicts. Thus, it is crucial to provide players with good explanations as an introduction to the game, and good facilitation throughout the gameplay, helping them focus on achieving goals and dealing with emotions in a fun way. Likewise, debriefing must address pain points and aha moments, link the game experience with reality, and help participants achieve objectives beyond the game session.

Finally, data collection during and after game sessions can be challenging for facilitators and evaluators when they are immersed in game sessions with participants. We try to address this by assigning roles. However, too big a group of "observers" can make participants uncomfortable.

We close by acknowledging the relevance of developments in Artificial Intelligence (AI) to game design. AI's powerful synthesis and pattern recognition capabilities, combined with the new feature of Natural Language Processing, provides new opportunities for serious game application (see Box 31.3). Using AI can save time, and allows us to focus on the unique value of the games as human interactions. The things that (so far) cannot be replaced by AI are: facilitating human sensemaking about complex and wicked problems (Edwards, 2023), experiencing individual and collective interactions driven by bounded rationality, seeing problems nested in a broader context, developing ideas about causal mechanisms, questioning the status quo, and exploring counterfactual conjectures of "what if" (Chomsky et al., 2023).

There is a word of warning, however: AI neural network learning models are a "black box." They could reproduce biases in content and language. Thus, program evaluators and game participants should stay alert for power imbalances and biases often hidden in systems recreated with AI.

BOX 31.3 AI: THE GAME CHANGER FOR SERIOUS GAMES?

DESIGN STAGE

Content Generation: AI can assist in generating game content and narratives by analyzing vast amounts of data, including academic resources and historical data. Evaluators can feed AI with program-specific data to create realistic and relevant scenarios. This reduces the manual effort required to design content and align the game with evaluators' goals.

Automated Playtesting and Bug Detection: AI can streamline playtesting by simulating various player behaviors and identifying potential bugs or issues. This reduces development time and ensures a smoother gaming experience for users, especially in digital games.

APPLICATION STAGE

Personalization and Adaptation: AI algorithms can analyze individual player behavior, preferences, and learning patterns, and help to ensure the game adapts to the unique profile of players, making it more engaging and effective at delivering educational content. Moreover, AI can adjust the game's difficulty level in real-time, providing appropriate challenges to keep players motivated and in the flow state.

Natural Language Processing (NLP): NLP capabilities can enhance communication within the game. Players can interact with AI-driven virtual characters using natural language, enabling more immersive and realistic experiences. Additionally, NLP can be utilized by program evaluators to provide real-time feedback, explanations, and assistance to players.

Prompts for lateral thinking: As reported by *Wired* magazine, AI AlphaGo's games against Lee Sedol produced moves that were unintuitive to human experts, but proved pivotal in shaping the outcome of a game (Metz, 2016). Thus, AI can nudge human players to rethink their strategies and even redefine the game.

POST-GAME STAGE

Player Assessment and Progress Tracking: For digital games, AI can continuously monitor player performance, collecting data on their decisions, interactions, and learning outcomes. These data can then be analyzed to evaluate the players' progress, identify areas of improvement, and provide insightful reports to evaluators. For boardgame sessions, AI can analyze players' decision record, evaluators' notes, and debriefing session minutes, to find patterns and derive valuable insights from individual sessions and comparisons across game sessions.

Source: Authors' own work based on conversation with ChatGPT-4 (prompt: "How can AI support serious games design process for public institutions and evaluation?").

CONCLUSION

In this chapter, we discussed the prospects of using serious games in evaluation. We demonstrated how games can be used to address many challenges faced in modern evaluation practice because they provide: a laboratory of decision-making, a safe space for experimentation, a multilogue that engages stakeholders, a time machine for anticipating effects, and a learning machines to transfer evaluative insights.

We organized the useful application of games in evaluation around two dimensions: the purpose of the evaluative inquiry, and the object of specific evaluation research. That organization structured the rich universe of analog and digital games into four major areas of applications: (1) assessing behaviors of individuals and groups, (2) adjusting behaviors of individuals and groups, (3) crash-testing mechanisms of institutional and collective behaviors, and (4) exploring complex systems.

We want to close this chapter with suggestions for a future and emerging research and practice agenda. We showed games as highly versatile tools that can be used for various purposes and applied to various policy topics. Looking at the broader landscape of public policies, there are two issues that, in our opinion, should, in particular, attract a shared effort of evaluators and game designers in the upcoming years.

The first issue is the resilience of local communities and societies. The global environmental crisis and its consequences require us to prepare for the unexpected and strive for transformation. Evaluating resilience means looking beyond direct emergency responses to explore sustainability questions, promote adaptability to unexpected systemic shocks, run stress tests, and question mainstream theories and assumptions underlying current programs, policies, and institutional settings. Those exercises could and should have different sectoral focuses (e.g., energy, food) and levels of governance.

Games can be employed to help evaluators think about and test for the unexpected. Games from areas three (crash-testing mechanisms) and four (exploring systems) discussed in this chapter could be especially valuable, as they can bring together complex socio-economic dynamics, behaviors driven by policy actors' bounded rationality, and rules of natural processes. For example, deliberative role-play and matrix game systems (Curry et al., 2018) allow for open conversations, argument-building, and premises-testing among players. Advances in data science can complement the use of games for evaluating resilience (e.g., big data on cities), and in modern technologies (in particular Augmented Reality). IT enhancements make game experiences in decision rooms more realistic, and support meaningful data-driven conversations around the game boards.

The second major issue is the challenge of supporting evidence-based policy in the contemporary post-truth context of alternative "facts," misinformation, and fake news (Marra et al., 2024). The post-truth era has made the jobs of public policy evaluators highly challenging. The abundance of misinformation erodes a factual base for diagnosing problems and designing policy solutions; and it obstructs implementation, by limiting the basis for consensus and coalition-building among stakeholders and the general public (Kavanagh & Rich, 2018). As a result, we may end up with policies that are designed based on how they make addressees feel, rather than how well they work in reality to address real problems. Thus, building respect for evidence among public program stakeholders is of primary importance.

For this challenge, evaluators could employ games to assess and adjust behaviors. For example, games can be used to test policy stakeholders for their immunity to fake news, educate users about the types and roles of evidence, show the value of data-driven decisions, and raise awareness among decision-makers about their own biases and gaps in reasoning. In short, games can build and test the capacity for evidence-based decision-making among individuals and teams involved in programs and policies. Furthermore, games can be used to transfer evaluation findings to the public, integrating facts and evidence into experiential game sessions to make them more digestible for stakeholders.

In order to use games effectively on those two topics and other policy issues that could emerge on future policy agendas, evaluators need to develop a shared conceptual language with game designers. In this chapter, we have provided a framework that could facilitate such communication. However, we see it only as a starting point for more in-depth conversations that could emerge during future joint undertakings by evaluators and game designers. A shared conceptual language will have value beyond practicalities. The language of games provides a structure for very straightforward questions like: How are the resources distributed among

players? Who sets the rules, and according to what underlying paradigm? What group of players are marginalized or even excluded from being at the table? Looking at evaluated programs and policy issues through the lenses of game language could advance often delicate conversations about inclusion, equity, and multicultural validity.

A decade ago, an author from the community of gamers pointed out that "reality is broken" (McGonigal, 2011), showing at the same time that games have a lot to teach us, in a non-threatening way, about how to make our lives better. We hope this chapter will convince evaluators to explore partnerships with the gaming community, to improve policies for addressing socio-economic challenges, and ultimately make the world a better place.

FUNDING INFORMATION

This chapter is a part of the research funded by the National Science Centre, Poland (grant number 2021/43/B/HS5/01935).

NOTE

1. Psychologists Gobet, de Voogt, and Retschitzki made the most original point (2004). They assert that games can be for decision and cognitive studies, similar to fruit flies for biology: a model organism used as a powerful means to gain knowledge that generalizes to a larger number of organisms. It is worth noting, however, that they made this point about board games. Unlike video games, board games have simplified rules presented to the players up front.

REFERENCES

Abt, C. (1987). *Serious Games*. University Press of America.
Adams, E. (2009). *Fundamentals of Game Design* (2nd ed.). New Riders.
Awan, T. M., Aziz, M., Sharif, A., Raza Ch, T., Jasam, T., & Alvi, Y. (2022). Fake news during the pandemic times: A systematic literature review using PRISMA. *Open Information Science*, *6*, 49–60.
Bamberger, J. M., Vaessen, J. L., & Raimondo, E. (2015). *Dealing with Complexity in Development Evaluation: A Practical Approach*. Sage.
Barton, M., Symborski, C., Quinn, M., Morewedge, C. K., Kassam, K. S., & Korris, J. H. (2016). The use of theory in designing a serious game for the reduction of cognitive biases. *Transactions of the Digital Games Research Association*, *2*(3), 61–87.
Baumgartner, F. R., & Jones, B. D. (2005). *The Politics of Attention: How Government Prioritizes Problems*. University of Chicago Press.
Beavis, C. (2017). Serious gaming. In K. Peppler (Ed.), *The SAGE Encyclopedia of Out-of-School Learning* (pp. 697–698). Sage Publications.
Blackett, P., FitzHerbert, S., Luttrell, J., Hopmans, T., Lawrence, H., & Colliar, J. (2022). Marae-opoly: Supporting localised Maori climate adaptation decisions with serious games in Aotearoa New Zealand. *Sustainability Science*, *17*(2), 415–431.
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*, 178–192.
Brathwaite, B., & Schreiber, I. (2008). *Challenges for Game Designers*. Cengage Learning.
Chomsky, N., Roberts, I., & Watumull, J. (2023). The false promise of ChatGPT. *New York Times*. Retrieved 2023-05-28.

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.
Costikyan, G. (2013). *Uncertainty in Games*. MIT Press.
Cousins, B. J., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, *80*, 5–23.
Csikszentmihalyi, M. (2008). *Flow: The Psychology of Optimal Experience*. Harper Perennial.
Curry, J., Engle, C., & Perla, P. (Eds.). (2018). *Matrix Games Handbook: Professional Applications from Education to Analysis and Wargaming*. The History of Wargaming Project. http://www.wargaming.co.
Dixson-Decleve, S., Gaffney, O., Ghosh, J., Randers, J., Rockstrom, J., & Stoknes, P. E. (2022). *Earth for All: A Survival Guide for Humanity*. New Society Publishers.
Djaouti, D., Alvarez, J., Jessel, J.-P., & Rampnoux, O. (2011). Origins of serious games. In M. Ma, A. Oikonomou, & L. Jain (Eds.), *Serious Games and Edutainment Applications* (pp. 25–43). Springer.
Donaldson, S., Patton, M., Fetterman, D., & Scriven, M. (2010). The 2009 Claremont debates: The promise and pitfalls of utilization-focused and empowerment evaluation. *Journal of MultiDisciplinary Evaluation*, *6*(13), 15–57.
Dörner, R., Göbel, S., Effelsberg, W., & Wiemeyer, J. (Eds.). (2016). *Serious Games: Foundations, Concepts, and Practice*. Springer.
Dryzek, J., & Pickering, J. (2019). *The Politics of the Anthropocene*. Oxford University Press.
Duke, R. (2011). Origin and evolution of policy simulation: A personal journey. *Simulation & Gaming*, *42*(3), 342–358.
Duke, R. (2014). *Gaming: The Future's Language*. W. Bertelsmann Verlag.
Duke, R., & Geurts, J. L. A. (2004). *Policy Games for Strategic Management: Pathways into the Unknown*. Dutch University Press.
Edwards, P. (2023). Serious games as an adaptive governance method. In S. Juhola (Ed.), *Handbook on Adaptive Governance* (pp. 115–125). Edward Elgar Publishing.
Elias, G. S., Garfield, R., & Gutschera, K. R. (2012). *Characteristics of Games*. MIT Press.
Esty, D. (Ed.). (2019). *A Better Planet: 40 Big Ideas for Sustainable Future*. Yale University 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.
Fullerton, T. (2019). *Game Design Workshop: A Playcentric Approach to Creating Innovative Games* (4th ed.). CRC Press.
Gobet, F., de Voogt, A., & Retschitzki, J. (2004). *Moves in Mind: The Psychology of Board Games*. Psychology Press.
Hiwiller, Z. (2016). *Players Making Decisions: Game Design Essentials and the Art of Understanding Your Players*. New Riders.
Huizinga, J. (1949). *Homo Ludens: A Study of the Play Element in Culture*. Routledge.
Kapp, K. M. (2012). *The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education*. Pfeiffer.
Kavanagh, J., & Rich, M. D. (2018). *Truth Decay: An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life*. RAND Corporation.
Klabbers, J. (2009). *The Magic Circle: Principles of Gaming & Simulation*. Sense Publishers.
Lasswell, H. D. (1951). The policy orientation. In D. Lerner & H. D. Lasswell (Eds.), *The Policy Sciences: Recent Developments in Scope and Method*. Stanford University Press.
LeBlanc, M. (2004). Mechanics, dynamics, aesthetics: A formal approach to game design. http://algorithmancy.8kindsoffun.com/MDAnwu.ppt.
Marra, M., Olejniczak, K., Paulson, A. (Eds.). (2024). *Evaluation in the Post-Truth World*. Routledge.
Mathison, S. (Ed.). (2005). *Encyclopedia of Evaluation*. Sage.
McCreight, R. (2019). *An Introduction to Emergency Exercise Design and Evaluation* (3rd ed.). Bernan Press.
McGonigal, J. (2011). *Reality Is Broken: Why Games Make Us Better and How They Can Change the World*. Penguin.
McGrady, E., & Curry, J. (2021). *Roll to Save: Gaming Disease Response. How to construct games in support of public health professionals*. The History of Wargaming Project. http://www.wargaming.co.

Meier, S. (2020). *Sid Meier's Memoir! A Life in Computer Games*. W. W. Norton.
Meijer, S. (2015). The power of sponges: Comparing high-tech and low-tech gaming for innovation. *Simulation and Gaming*, *46*(5), 512–535.
Metaari (2021). *New Metaari forecast for worldwide 2020–2025 game-based learning market shows North America sales will soon outstrip Asia Pacific as game revenues in China slow*. https://seriousplaywire.com/metaari-game-based-learning/.
Metz, C. (2016, March 16). In two moves, AlphaGo and Lee Sedol redefined the future. *Wired*.
Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129–140.
Olejniczak, K. (2017). The game of knowledge brokering: A new method for increasing evaluation use. *American Journal of Evaluation*, *38*(4), 554–576.
Olejniczak, K., Newcomer, K., & Meijer, S. (2020). Advancing evaluation practice with serious games. *American Journal of Evaluation*, *41*(3), 339–366.
Olejniczak, K., Raimondo, E., & Kupiec, T. (2016). Evaluation units as knowledge brokers: Testing and calibrating an innovative framework. *Evaluation*, *22*(2), 168–189.
Olejniczak, K., Wolański, M., & Widawski, I. (2018). Regulation crash-test: Applying serious games to policy design. *Policy Design and Practice*, *1*(3), 194–214.
Ostrom, E. (2005). *Understanding Institutional Diversity*. Princeton University Press.
Palenberg, M., & Paulson, A. (Eds.). (2020). *The Realpolitik of Evaluation: Why Demand and Supply Rarely Intersect*. Routledge.
Patton, M. Q. (2010). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. Guilford Press.
Patton, M. Q. (2019). *Blue Marble Evaluation*. Guilford Press.
Pawson, R. (2013). *The Science of Evaluation: A Realist Manifesto*. Sage.
Peters, B. G. (2018). *Policy Problems and Policy Design*. Edward Elgar Publishing.
Roberts, J. D., Chung, G. K. W. K., & Parks, C. B. (2016). Supporting children's progress through the PBS KIDS learning analytics platform. *Journal of Children and Media*, *10*(2), 257–266.
Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1), 65.
Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). *Evaluation: A Systematic Approach* (6th ed.). Sage.
Salen, K., & Zimmerman, E. (2004). *Rules of Play: Game Design Fundamentals*. MIT Press.
Schrier, K. (2021). *We the Gamers. How Games Teach Ethics and Civics*. Oxford University Press.
Scriven, M. (1991). *Evaluation Thesaurus* (4th ed.). Sage.
Sellers, M. (2017). *Advanced Game Design: A Systems Approach*. Addison-Wesley.
Service, O., Hallsworth, M., Halpern, D., Algate, F., Gallagher, R., Nguyen, S., Ruda, S., & Sanders, M. (2014). *EAST: Four Simple Ways to Apply Behavioural Insights*. Behavioural Insights Team.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin.
Shafir, E. (2013). Introduction. In E. Shafir (Ed.), *The Behavioral Foundations of Public Policy* (pp. 1–9). Princeton University Press.
Shute, V., & Sun, C. (2020). Games for assessment. In J. Plass, R. Mayer, & B. Homer (Eds.), *Handbook of Game-Based Learning* (pp. 491–512). MIT Press.
Smith, N. (2021, August 11). A Yale doctor is using a video game to fight the opioid crisis. *The Washington Post*.
Stokes, B. (2020). *Locally Played: Real-World Games for Stronger Places and Communities*. MIT Press.
UN News (2020, April 13). During this coronavirus pandemic, 'fake news' is putting lives at risk. UNESCO – *UN News*. https://news.un.org/en/story/2020/04/1061592.
Weiss, C. H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice*, *9*(1), 5–19.
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.
Woods, S. (2012). *Eurogames: The Design, Culture and Play of Modern European Board Games*. McFarland & Company.

Zhonggen, Y. (2019). A meta-analysis of use of serious games in education over a decade. *International Journal of Computer Games Technology*, *2019*, 1–8.

APPENDIX

Table 31A.1 Key concepts for communicating with game designers

Element of the ABC framework	Detailed definition and description
GAME PURPOSE	The intent for which the game is used as a method for public policy. It can be (1) assessing behaviors, (2) adjusting behaviors, (3) crash-testing mechanisms, or (4) exploring systems. In short, who will use the game and for what.
ARCHITECTURE	All elements that are used to design a fully-fledged game.
Aesthetics	The narrative story of the game, its form, and graphic design that tries to build the ambiance of the game story.
Objective in the game	The objective players pursue within the game. This is the ultimate set of conditions for winning the game.
Actors	Types of players, their roles, and their starting characteristics within the game. Also includes non-player characters (NPC), that is, actors controlled by the game with a predetermined set of behaviors.
Action arenas	Arenas in which players perform their actions. These could be boards, created worlds, spaces, etc. There might be several different, overlaying arenas, each containing additional space for decision-making or resource management.
Resources	The level of time and items at the players' disposal. It can change over the course of the game due to players' choices or rules enforcement.
Rules	A shared understanding by the participants about enforced prescriptions concerning what actions (or outcomes) are required, prohibited, or permitted.
Mechanisms	Interdependencies and causal connections between elements of the systems, especially actors' behaviors, resources, and rules.
Game loop	A sequence of actions forced by game rules and performed by players during the game.
Decision situations	Specific key decisions and trade-offs players are confronted with during the game.
Drivers (also known as mechanics)	Key dynamics introduced to the game in order to challenge players, and drive their behaviors and progress in the game. These are such things as competition, ticking clock, exploration, etc.
FEEDBACK	The return information about players' performance (also called "adjudication"). It is the outcome of the process and actions of the players, and the way in which players learn about their progress in achieving their objectives. A scoring system is required to indicate to what extent players' strategies are successful.
BEHAVIORS	Behaviors observed among players during their interaction with the game architecture.
Pain points	Struggles, frustrations, challenges, and difficult moments that players experience during the game session.
Aha moments	Revelations, discoveries, and enlightenment moments that players experience as the game progresses.
Strategies	A prevailing pattern of choices and responses players slowly develop and pursue over the gameplay.
DEBRIEFING	A guided reflection on the mechanisms experienced by players, and translation of those observations into reality. It serves the purpose of turning the experiences of players into knowledge and learning beyond the game.
CHANGE	The change observed after the game that can be causally linked with the completed game session.
Players' takeaways	Players' main memories and takeaways from the game (as reported in their own words). This includes the emotional recollection that remains after the game session.

Element of the ABC framework	Detailed definition and description
Participants' gains	Change (declared or observed) in players' specific skills, knowledge, behaviors, trust level, social network, and/or in their environment, due to participation in the game session.
Game principal gains*	Change in the principal's knowledge and understanding of the policy situation due to the execution of the game.

Note: * A game principal is a person or an organization that applies the game to the public policy issue. In the case of this chapter, principals are most often evaluators and the organization engaging them in the evaluation study.
Source: Developed by authors based on: Elias et al. (2012); Hiwiller (2016); Ostrom (2005); and Salen & Zimmerman (2004).