Advancing Evaluation Practice With Serious Games

Karol Olejniczak, Kathryn E. Newcomer, and Sebastiaan A. Meijer

Abstract
Evaluation professionals need to be nimble and innovative in their approaches in order to be relevant and provide useful evidence to decision-makers, stakeholders, and society in the crowded public policy landscape. In this article, we offer serious games as a method that can be employed by evaluators to address three persisting challenges in current evaluation practice: inclusion of stakeholders, understanding of causal mechanisms, and utilization of evaluation findings. We provide a framework that distinguishes among games along two crucial aspects of evaluation inquiry - its function and the nature of the evaluand. We offer examples of successfully implemented games in each set of the four arenas we delineate: teaching knowables, testing retention, crash-testing mechanisms, and exploring systems. We explain how games can be employed to promote learning about and among stakeholders, and to collect valuable intelligence about the operations of programs and policies.

Keywords
serious games, inclusion of stakeholders, causal mechanisms, evaluation use, evaluation methods

Evaluation needs to be nimble and innovative in order to be relevant and provide useful insights for decision makers, stakeholders, and society in the crowded public policy landscape. This point has been emphasized recently in different professional platforms by leaders of associations, evaluation theorists, and the general evaluation community (e.g., see Alkin & King, 2017; Australasian Evaluation Society, 2018; European Evaluation Society, 2018; Newcomer, 2017). The growing use of serious games in business and research presents opportunities for evaluation practice as they have not been widely employed in the public policy context (Abt, 1969/1987; Febey & Coyne, 2007).

1 Institute of Social Sciences, SWPS University of Social Sciences and Humanities, Warsaw, Poland
2 The Trachtenberg School of Public Policy and Public Administration, George Washington University, Washington, DC, USA
3 Department of Biomedical Engineering and Health Systems, KTH Royal Institute of Technology, Flemingsberg, Stockholm, Sweden

Corresponding Author:
Karol Olejniczak, Institute of Social Sciences, SWPS University of Social Sciences and Humanities, Chodakowska Street 19/31, 03-815 Warsaw, Poland.
Email: karol.olejniczak@swps.edu.pl
Public Policy and Challenges for Evaluation

At its core, public policy is about enabling, and often requiring, change in behaviors among the public and institutions. Policymaking requires designing and implementing triggers for change, often with inadequate information, in a complex environment (Bardach, 2006; Lasswell, 1951; Wildavsky, 2018). Evaluation entails assessing the worth and merit of the changes and providing insights to decision makers and stakeholders on which interventions work for whom. Public policies and programs are evaluated to assess their effectiveness and efficiency as well as the extent to which they contribute to social betterment and justice (Chelimsky, 2006; Dahler-Larsen, 2012; Henry & Mark 2003; House, 2017).

Three challenges emerge of primary importance for evaluation to be relevant and useful: inclusion, understanding, and utilization. Inclusion means including stakeholders, especially the targeted impactees in deliberations about both the design and evaluation of policies and programs. Effective inclusion entails recognizing and overcoming biases; fostering race, class, and gender sensitivity; and incorporating the perspectives of marginalized groups into policy and program design and evaluation (Young, 2000). Intentional inclusion is foundational for democracy, and it may help reduce policy makers’ confirmation and framing biases (Hallsworth et al., 2018). Securing consensus of stakeholders helps to provide stronger legitimization for actions and helps in implementation since coalitions of diverse actors are typically needed to develop and deliver solutions to complex social problems (Christensen, 2006).

The challenge of understanding means policy makers and evaluators need to unpack complexity and provide insights about the problem addressed and about the causal mechanisms needed to produce the desired change (Astbury & Leeuw, 2010). By its very nature, public policymaking is a trial and error process, and evaluation can make this process more evidence-informed by generating knowledge on what works for whom and in what context (Pawson, 2013). Understanding and analyzing complexity in both problems and solutions is vital for designing effective policy tools and replicating promising public interventions.

The utilization challenge entails bringing evidence (including that produced by evaluations) to policy practitioners to inform their decision-making process (Alkin & King, 2017; Nutley et al., 2007). Individual and organizational policy actors absorb information and learn in complex, nonlinear ways (Leeuw et al., 1994). Despite the well-intended efforts of evaluators to be utilization focused in their work, the transfer of knowledge from studies to decisions continues to be elusive (King & Alkin, 2019; Patton, 2018). More effective translation of findings to action could substantially strengthen policymaking and ultimately the effectiveness of public interventions.

Promising Advantages of Games for the Public Policy Arena

To address the persisting challenges to inclusion, understanding, and utilization, we suggest that evaluators could employ serious games in their work to help evaluation practice become more relevant and impactful. At least four general characteristics of games highlight their potential utility for use in evaluation practice: their universal language, flexibility in exploring uncertainty and complexity, capacity to facilitate learning, and the opportunity they afford for timely collection of relevant data.

Games offer a universal language of communication that can be used to help evaluators better achieve inclusion and utilization. Cultural studies indicate that play and games have permeated human civilizations, influencing informal and formal social structures and institutions (Caillois, 2001; Huizinga, 1949). To put it simply, games are affected by culture and also present a way to learn about culture (Ivory, 2018).
Furthermore, recent data indicate that digital games have become ubiquitous in our societies. For example, 64% of U.S. households own a device that they use to play video games, 60% of Americans play video games daily, 45% of U.S. gamers are women, and the average age of gamers is 34 years (Entertainment Software Association, 2018). Thus, games present a lingua franca to evaluators to help explain cultures in an inclusive way and a way to find common ground among diverse stakeholders. In addition, games use both a language of theory (games as testing hypotheses) and a language of practice (games as problem-solving activity). This logic is compelling for both researchers and practitioners, providing an opportunity for better knowledge transfer.

Games offer a nonthreatening way to explore and explain the complexity and uncertainty that are inherent characteristics of most policy issues. Even common games have been used as powerful metaphors to explain complex socioeconomic processes. One of the best-known examples is the game of chess. In the mid-13th century, Jacobus de Cesolli (a Dominican friar) used a chess metaphor to explain a significant shift in European governance. By the 15th century, his book was surpassed in popularity only by the Bible (Johnson, 2016).

An improved understanding of complex mechanisms can also be obtained by the participants as they experience the consequences of their actions. For example, an original version of Monopoly called the Landlord’s Game was designed in 1903 to make the public aware of the dangers of capitalist approaches to land taxes and property renting. The game’s commercial version was sadly stripped of this educational aspect (Pilon, 2015).

Games are appealing because “they allow us to explore uncertainty, a fundamental problem we grapple with every day, in a non-threatening way” (Costikyan, 2013, p. 13). In the public policy context, games present inexpensive opportunities to experiment with different strategies to address wicked problems. For example, the Rubber Windmill game was used by the British Office for Public Management to explore unintended consequences of restructuring the UK’s National Healthcare System (Duke & Geurts, 2004). This appealing feature of games is enabled by the “magic circle”—the fact that game’s space has its boundaries and rules does not intervene in reality and can be stopped without harm (Huizinga, 1949; Klabbers, 2009).

Games are also “learning machines” (Beavis, 2017, pp. 698–699). Learning through games is emotionally engaging, social, experiential, context situated, problem based, risk free, and accompanied by immediate feedback (Connolly et al., 2012; Iten & Petko, 2016). Those attributes align games closely with human mechanisms of learning for individuals, teams, and organizations (Kapp, 2012). Games can be used with diverse stakeholders to allow them to share different types of knowledge: skills based (technical and motor skills), cognitive (conceptual, procedural, and strategic knowledge), and affective (beliefs, values, and attitudes; Connolly et al., 2012). For example, games have been employed in the training of medical personnel (Graafland et al., 2012), to teach principles and raise awareness about flooding issues (Rebolledo-Mendez et al., 2009), and to educate about epidemic mechanisms and prevention (Pandemic series of popular board games). Conclusions from research on the relative effectiveness of games, versus other modes of instruction, have been mixed, but there has been a growing body of evidence on the efficacy of the game format for promoting behavioral and attitudinal change (Boyle et al., 2016; DeSmet et al., 2014; Giessen, 2015; Girard et al., 2013; Wang et al., 2016; Wouters et al., 2013; Zhonggen, 2019). Addressing motivations to encourage learning could be especially useful for boosting both understanding and utilization, while the socialization aspect of games could contribute to inclusion.

Finally, games present a promising method to capture timely data on how people actually behave, not what they declare. Measurement can be taken pregame, during game, and postgame. The spectrum of possibilities for collecting various data and diverse modes of collection include experiments with biometrics, psychometrics, path tracking, time stamping, ethnographic observation of
behaviors in games, interviews, focus groups, Q methodology, and scenario analysis (for an extensive overview of possibilities, see Mayer et al., 2014).

The value of games as a sound research method has been recognized in the social sciences (de Vaus, 2006) and in the public policy literature (Dunn, 2017, p. 18) because games allow observation of the effects of “interventions” in a controlled context. For example, Elinor Ostrom, Nobel laureate in economics, used games to test how communities across the world might effectively govern the commons and protect precious environmental resources (Ostrom et al., 1994).

Game processes have emerged as a useful data collection method in the digital age, as illustrated by the move to gamification in surveying (Salganik, 2018, p. 115). Relatedly, in the worlds of complexity science and engineering systems, games have been used to evaluate the contextual embeddedness of potential solutions (Grogan & Meijer, 2017; Raghothama & Meijer, 2018). As Verhagen et al. (2017, pp. 3–4) point out: “Given that in social sciences or societal challenges experiments in the laboratory or the real world are hard or even impossible to perform, games can be a new way to study how people interact, learn and collaborate in managing complex situations and problems.”

**Goal and Structure of the Article**

Game designers in related fields have been busy and prolific, and the number and variety of games is overwhelming. However, there have been only sporadic attempts to develop useful typologies to organize thinking about games relevant to public policymaking (Mayer et al., 2016; Sawyer & Prince, 2008), and there have been no attempts to propose conceptual frames about games for evaluation practice. Despite their potential, the use of games in evaluation has been rare. There is a dearth of literature on the use of games for policy and program evaluation. The isolated examples focus mainly on training applications, leaving out consideration of how games might be employed to improve inclusion and exploration of complexity in evaluation practice (Febey & Coyne, 2007; Olejniczak, 2017).

Our goal here is to draw from experience with serious games to promote games suitable and useful for advancing evaluation practice. We offer a framework to organize games around evaluation purposes and basic characteristics of the evaluand. We developed our framework based upon a systematic literature review on the use of serious games in the public policy arena (databases of Scopus and Web of Science, books query in 2017 and 2019) and our own experience with employing games in evaluation.

In the remainder of this article, we first present a brief overview of the origins, evolution, and nature of serious games practice. We propose a basic definition of serious games adapted to evaluation craft and offer a framework for conceptualizing how and when games may be employed in evaluation practice. Second, we illustrate each of the four areas of applications delineated in our framework with practical examples from the public policy arena. Third, we elaborate more on the three challenges to evaluation practice that games may help address, that is, promoting inclusion, understanding, and utilization, and we discuss the utility of identified game applications for addressing those challenges. We close by summing up the opportunities for applying serious games in evaluation practice and suggesting directions for future research. We hope that our overview of the field of serious games with applicability to evaluation practice will inspire our evaluation community and create opportunities for further knowledge exchange and synergies between the fields of evaluation and simulation/gaming.
The Origins and Nature of Serious Games

Serious games are nested in a broader ecosystem of gaming/simulations that is highly polymorphic and encompasses: cultural studies with anthropological and ethnographic approaches, social science research including education and psychology, business management, computing, and mathematics. And the academic literature on serious games has been growing in the last decade (see Figure 1). We provide here a brief overview of the more influential scholarship affecting design and practice with serious games.

The origins of serious gaming practice can be traced from early forms of war-themed games (chaturanga, chess, checkers), through the 17th- and 18th-century military games, to advanced World War II operation simulations, followed by military and policymaking games in the 1950s (Duke, 1974/2014; Wilkinson, 2016). However, the term “serious game” was introduced in 1969 by Clark Abt to refer to “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” (1969/1987, p. 9). Abt suggested that games could be used to motivate students by demonstrating to them the relevance to the real world of the abstractions they are taught. He illustrated the educational value of games with examples of both analog and digital games that were employed to teach children from disadvantaged groups, provide teenagers with occupational choice training, and allow adult decision makers to experiment with different strategies for solving problems.

The use of serious games to foster collaborative problem-solving was advanced by Richard Duke (1974/2014). Given the growing complexity of the modern world, Duke (1974/2014) noted that games offer symbolic maps of various multidimensional phenomena; the game players gain experience with the phenomena, and learn through discussion with various stakeholders, and thus gain a
gestalt understanding of the complex systems. Duke (2011) offered a number of serious games, including his Metro project—a game that helped the Lansing Michigan City Council in the United States to simulate budgeting processes showing policy trade-offs, resource constraints, and raising empathy for social issues. He initially used the term “gaming.” However, in later publications, he used the term “policy games” interchangeably with “serious games” (Duke & Geurts, 2004). Duke’s work was instrumental in the development of the professional association for game developers—the International Simulation and Gaming Association.

The explosion of computer games in the 1980s pushed serious gaming toward digital forms and increased the use of games in teaching. Games were employed to support instruction in elementary schools, as well as higher education, vocational training, and collaborative workplace training (Dörner et al., 2016). Outside of the classroom, the use of games increased in training simulations, environmental awareness exercises, and basic math instruction through the use of digital puzzles. In 2011, Djaouti et al. reviewed the leading platform of serious games and found that in the pre-2002 period, of 953 games, over 65% were used for educational purposes (p. 36).

In 2002, the potential for the use of serious games in the public policy arena was promoted in a white paper by Sawyer and Rejeski (2002). They suggested that digital commercial game practices and technology could be utilized by public organizations. Sawyer and Rejeski did not provide specific definitions or exemplars, but they broadly promoted the use of computerized game/game industry resources whose chief mission was not entertainment. The report triggered further initiatives, including the Woodrow Wilson International Center for Scholars Serious Games Initiative. As Djaouti et al. (2011) point out, the Sawyer and Rejeski’s white paper came out about the time of the release of the U.S. Army’s highly successful game used in basic training which was used as a tool for recruitment and promoted broader public awareness of serious games.

The recent increase in the number of publications discussing serious games has been primarily in the following areas: computer science and engineering, with a focus on the development of games, and educational research, psychology, and various health sciences, with focus on applications of games and their outcomes (Ciftci, 2018; see Figure 2). There have been some very innovative

Figure 2. Distribution of serious games publications among research areas. Source. Web of Science database (accessed August 20, 2019).
applications in health care, including Re-Mission, which is used by cancer patients who control a nanobot to fight cancer infections and change their attitude toward chemotherapy treatments, and SnowWorld, which is used to distract burn victims from pain during wound treatment by immersing them in a virtual world snowball fight (Dörner et al., 2016).

The most prevalent definition of serious games in recent literature is as follows: digital games that do not have entertainment as their primary purpose but focus on training, learning, and behavioral change (in both formal and informal learning contexts; Beavis 2017; Dörner et al., 2016; Michael & Chen, 2006; Ricciardi & De Paolis, 2014). However, there are still differences across fields in definitions employed. Sometimes serious games are not characterized by the intention of the developer but by the intention of the player (Dörner et al., 2016) or by the topic they address, for example, cancer, genocide in Darfur (Fullerton, 2019). Some authors further distinguish between simulations that are realistic and support first-order learning skills and serious games that are more abstract and are designed to improve second (higher)-order learning and thinking (Charsky, 2010).

Some confusion about serious games has arisen due to the gamification movement in which game mechanics and elements, for example, mission, points, badges, and levels of achievement, are applied to nongame activities, for example, sales, office tasks, and class activities, in order to make these activities more engaging (Routledge, 2015). And recently, a related phenomenon has emerged—knowledge games (also known as GWAP—Games With a Purpose) that are employed to exploit crowdsourcing and engage citizens in scientific knowledge generation and collective problem-solving (Dörner et al., 2016, p. 6; Schrier, 2016). Two prominent examples of knowledge games are Foldit, that engages playersto fold the structures of selected proteins as perfectly as possible using tools provided in the game, and PHYLO, which is a game framework used to compare genomes and solve the Multiple Sequence Alignment problem.

Despite the increasing interest in the use of serious games, it is not a fully formed field of practice, and it lacks a unified theoretical basis and coherent methodology. The fluidity in the field reflects, in part, the large number of disciplines contributing to scholarship about games. The permeable space spurs innovation but also allows for “dilettante” research and developmental practices (Wilkinson, 2016, p. 28). Given the technology-driven evolution of serious games, current practice also tends to be narrower than the initial visions of serious games. It is focused on digital games and using games to improve individuals’ competencies, giving less attention to analog games designed to allow players to collectively address “wicked” public policy problems (Duke, 2011; Klabbers, 2018). In our view, analog games present potential opportunities for evaluation practitioners.

**Definition of Serious Games for Evaluation**

For the application of serious games in public policy and evaluation, we define serious games as analog or digital games used within a well-defined space, which have a clear primary purpose different from entertainment and intentionally transfer the game experience to reality. Both the purpose and way of transfer are determined by the game principal, who is a person or organization that applies the game to a public policy issue. What distinguishes games, and serious games in particular, from simple play is the presence of explicit rules (Dörner et al., 2016) and a clear purpose or goal (Caillois, 2001; Gray et al., 2010).

Our definition overcomes some limitations in contemporary literature on “serious gaming” in terms of forms and function. We include analog (tabletop) and digital (computer, smartphones) games and games that develop individual competences (education and training games), test behaviors in realistic settings (simulations), and facilitate collective dialogue and problem-solving
among stakeholders (policy games). Our definition of serious games also reflects the emphases given to gaming in the work of pioneers such as Abt (1969/1987) and Duke (1974/2014). We distinguish between serious games and gamification, as gamification tends to take place outside of a defined space and applies game concepts to market driven or individual objectives, such as making sales, walking up steps, and tracking the number of miles run (Kapp, 2012). We also distinguish serious games from entertainment games and stress the applied character of serious gaming. Players’ experience with games for entertainment stops with the game, but serious games focus on how the experience gained in the game is used to change behavior afterward (Wenzler, 2019). Finally, we highlight the role of the game principal and stress that there are a variety of potential principals, including researchers, policy designers, and intended policy or program beneficiaries.

A Framework for Application of Serious Games to Evaluation

In other arenas, typologies of games have been offered that differentiate different genres, medium used, game mechanics, and primary functions (Hartevelt, 2011; Mayer et al., 2016; Michael & Chen, 2006). However, none of those typologies aligns games with uses in public policy design, implementation, or evaluation. Therefore, we developed a framework explicitly focused on the utility of games in public sector evaluation and analysis. It is presented in Figure 3.

We chose two fundamental dimensions to distinguish among games that may advance evaluation practice: the primary function of evaluation inquiry and the nature of the evaluand.
Our first dimension ranges from testing to triggering change. This dimension describes the fundamental objective about the use of games to inform policy designers and/or intended policy or program beneficiaries. At the testing end of the continuum, the game is used as a research method to study a particular policy intervention, and those impacted by it, and measure change in their situation and behaviors. Evaluators are observers, while participants of the game and their behaviors are subjects of the observation and assessment. This option is aligned with a positivist approach to evaluation inquiry, where the researcher controls and manipulates the experimental factors and observes the reactions of the subjects.

At the other end of the continuum—triggering change—the game presents an intervention and is used as an instrument for generating change in cognition and behaviors among participants of the game. This application is more aligned with a constructivist, action research approach, where evaluators and participants are engaged in interactions and cocreate social change (Fetterman & Wandersman, 2005; Patton, 2010).

Clarifying a function of evaluation is also crucial for determining the audience that learns from the conducted evaluation inquiry. In the case of testing change, learning is directed primarily toward designers of the particular intervention who are not directly involved as participants in the game. In the case of triggering change, the participants in the game are actively learning. The game experience is designed to foster a shared understanding among those who engage in the game, as well as produce knowledge and skills among participants over the long term.

Our second dimension focuses on the nature of the evaluand and ranges from complicated to complex systems. An evaluand is the object of an evaluative inquiry. Evaluands are typically policy instruments like projects, programs, and regulations (Mathison, 2005). This dimension describes the fundamental assumption about the level of structure, order, and predictability that exists in the system under inquiry. A complicated system has a limited number of elements, with relatively clear cause-and-effect relationships. Complex systems involve a larger number of interacting elements whose interactions are likely to be nonlinear and where small changes may produce significant consequences. In more complex systems, we may observe emergent patterns and causal links among emergent elements; a whole system may evolve in response to environmental factors or other systems (Snowden & Boone, 2007; Williams & Hummelbrunner, 2010).

Understanding the nature of the evaluand has critical implications for methods choices in evaluation. Complicated systems may allow establishing causal inference through the use of experiments and quasi-experiments and applying quantitative modeling to well-structured policy problems (a hard system thinking approach). On the other hand, complex systems require more exploratory approaches that draw upon different types of knowledge held by stakeholders and integrate the data collected (a soft system thinking approach; Duke & Geurts, 2004, pp. 148–150; Ramage & Shipp, 2009).

The two-dimensional framework distinguishes among four broad sets of applications that may be especially useful in the context of evaluation of public policies and programs: Teaching Knowables (complicated evaluands with a focus on triggering change), Testing Retention (complicated evaluands with a focus on testing), Crash-testing Mechanisms (complex evaluands with a focus on testing), and Exploring Systems (complex evaluands with a focus on collective problem-solving and triggering change). Initially, clarifying the focus (a type of evaluand) and the purpose (testing vs. triggering change) is especially critical for the intended users of the results, as well as for the game designers. Designers can modify games, so while we introduce examples of games that were designed for testing, modifications may be made to those games to make them suitable for promoting problem-solving and change in addition to simply assessing retention of knowledge.
How and When Might Serious Games Be Employed in Evaluation Practice?

In this section, we provide an overview of game applications aligned with our framework. For each of the four areas of application, we offer a brief explanation about the objective, followed by a description of an example that illustrates how this type of game has been applied. For each example, we explain the policy context, purpose, and audience of the game and its content, outcomes, and the potential adaptability of the game for other areas of application.

Teaching Knowables

In evaluations, we often try to educate stakeholders about principles, mechanisms, and effective strategies relevant in a particular policy issue or program context. Instead of explaining these concepts and dynamics through traditional formats, games can be used to help stakeholders learn through experiencing the relevant phenomena at work.

The first set of games focuses on teaching participants. The game masters, that is, those who actually run the games, must possess a well-structured understanding of the policy issue or process that is the subject of evaluative inquiry. They can then use a game to transfer this knowledge to the target group through engaging them in problem-based, experiential learning. Typically, game sessions consist of multiple game rounds interwoven with facilitated debriefings that allow participants to reflect on their experience and gradually build their understanding.

Exemplar 1: Knowledge Brokers

**Policy context, purpose, and audience.** Our example of the Knowledge Brokers game-based workshop illustrates how evaluators can use games to build evaluation capacity among staff of government units (Olejniczak, 2017). In 2015, evaluation units in the Polish national government started to search for ways to address the gap between the increasing production of evaluation studies and the limited use of their findings. A set of comparative studies had identified knowledge brokering as a promising practice for addressing the challenge of using evidence to inform policy design and implementation (Olejniczak et al., 2016). Knowledge Brokers are typically government staff who help policy makers and other potential evidence users develop the questions they would like addressed and then ensure that evaluation studies and data answer the knowledge needs. The challenge to address is how to teach busy staff in evaluation units the principles and skills of knowledge brokering.

**Game content.** A team of evaluators and game designers developed a game-based workshop to help participants develop knowledge brokering skills. A one-day session was designed for a group of up to 30 policy professionals, who work in teams of 5–6. The workshop encompassed three elements: the game, mini lectures, and debriefings. In the board game, teams manage evaluation units in specific geographic regions for 12 months. During that time, four different public interventions are implemented in their regions. At different stages of those interventions, different decision makers make requests for evaluations of the interventions. Players have to provide credible knowledge to each key user in a timely fashion and in an accessible way, despite having limited time and human resources.

The game rounds allow participants to experience the dynamics of the real work of a knowledge broker. Mini lectures provide participants with explanations of key factors that affect the effectiveness of their work, and the debriefings allow players to reflect on their strategies and ultimately transfer key learnings into the practice of their organizations. The objective is to facilitate learning among participants, and perhaps their teams, that they can employ in their current positions.
Outcomes of the game. Broadly speaking, participants of the workshop learn how to effectively use results of evaluation studies to inform decision makers. The five desired learning outcomes for participants are (1) identifying specific knowledge needs by matching types of researchable questions to the interests of different types of decision makers, (2) acquiring credible knowledge by matching appropriate research designs to the type of knowledge needed (the researchable questions), (3) applying effective methods to inform decision makers by matching forms and channels of communication with the typical preferences of that particular type of knowledge user, (4) strengthening policy arguments by combining findings from single studies into one knowledge stream, and (5) being aware that knowledge resulting from evaluation efforts is only one of the elements taken into account in the decision-making process.

The Knowledge Brokers workshop has been recognized as useful by a number of evaluation units that have applied it for capacity building among their staff and across organizational networks of their clients, for example, Czech and Polish national evaluation units, the European Bank for Reconstruction and Development, the U.S. National Science Foundation, and the National Public Health Agency of Canada. However, so far, no systematic study has been executed to measure retention of skills or the long-term organizational impact of this training.

Adaptability. A follow-up case of using the Knowledge Brokers game illustrates how a game focused on teaching knowables may be redesigned for system exploration purposes.

The Foundations for Evidence-Based Policymaking Act in 2018 (US Congress 115/4174) recently introduced a statutory requirement for all federal agencies to prepare evidence-building plans (so-called Learning Agendas). The plans are expected to be short documents that integrate different streams of evidence and learning activities around key questions important for agency mission and operations. In principle, Learning Agendas should be developed collaboratively with extensive involvement of stakeholders, facilitate agency’s evidence building, and improve internal decision-making (Baker, 2017; Nightingale et al., 2018). A key challenge facing agency leadership is to design a process that meaningfully engages both their agency’s staff and other stakeholders and produces a valuable map for building useful evidence.

Evaluators in the Polish national government and the evaluation unit in the U.S. National Science Foundation have used the Knowledge Brokers game to develop Learning Agendas for their agencies. Participants of the session use the key elements of the game (decision moments, types of knowledge users, knowledge needs, sources of information, and communication strategies) as building blocks to design learning agendas that are tailored to their strategy or program. Through use of the game, practitioners are able to (a) focus their research activities along timelines and decision moments, (b) target information needs of their key decision makers (knowledge users), and (c) integrate different knowledge sources (including evaluations) and learning methods to inform program decision-making. This game is a fully open system that allows practitioners to explore, discuss, and eventually design knowledge management around their specific public program or strategy.

Testing Retention
One of the traditional areas of evaluation is assessment of the knowledge, skills, and competencies of the stakeholders involved with a particular policy or program. Tests and surveys typically are used to test intended beneficiaries, perhaps better labeled as impactees. Games in this category can be used to evaluate participants’ real, not declarative competencies, skills, and knowledge by observing their behavior because they allow measurement in a dynamic, but still controlled environment. Game participants are exposed to specific situations, and their behaviors are observed and evaluated in line with the desired skills and competencies.
**Exemplar 2: ProRail—the Dutch railway administration**

**Policy context, purpose, and audience.** Our example is a simulation developed for ProRail, the Dutch railway administration, that was the game principal and needed expert guidance on a high-stakes scheduling challenge. ProRail was facing a major capacity issue on an important railway line between Amsterdam and the eastern region of the Netherlands. A key challenge was that a specific bridge needed to open to permit boats to pass, and the requirement by law was to open this bridge twice per hour for 5 min. This requirement of adding 10 min per hour to the timetable was problematic as the increasing demand placed on the railway exceeded the remaining 50 min per hour.

One policy change being considered was to elevate the tracks over many kilometers and even elevate a nearby railway station, for an estimated cost of about 500 million Euro. Another option raised was to allow train traffic controllers to adaptively open the bridge twice per hour, in response to the demand, instead of the planned pattern. This second solution could save massive investments in hardware infrastructure. However, it was deemed essential to ensure that the train traffic controllers would have the requisite skills to run a more adaptive system, and a game was developed to test the traffic controllers on these skills (Kortmann & Sehic, 2011).

**Game content.** The game consisted of a computerized simulation of train traffic flow, with interfaces to control the simulation that were largely similar to the human–machine interface of a real train traffic control workstation. The game was designed to be played by individual train traffic controllers with varying levels of experience. Skills in dynamic problem-solving (which went beyond what was covered in the existing operating instructions but were developed in a new training program) were tested by having participants operate in scenarios with variations in the train traffic patterns and contextual pressures from train drivers and train traffic controllers from neighboring areas, whose roles were played by the game masters.

**Outcomes of the game.** Two rounds of play were organized, with an improved version of the game (better interface, improved validity of the underlying simulation) in the second round (Meijer, 2015). Based on the observations of a large number of individual players, the conclusion was drawn that it is possible for a large portion of the train traffic controllers to open the bridge in a dynamic regime, thereby minimizing the effect of bridge openings on railway capacity. Experience was found to be an important predictor affecting the ability of participants to open the bridge at optimal intervals, with 10 years of experience emerging as the optimal experience level for the traffic controllers. Based on this game, the decision was made to not invest in elevating the railway tracks but to invest in improving the capacity of all traffic controllers to manage bridge operations more adaptively.

**Adaptability.** Subsequently, the modules that constituted the game were extended to include other operational roles as well, and the extended simulations have been used to explore additional issues in transportation systems in the Netherlands.

**Crash-Testing Mechanisms**

Successful policy solutions typically rely on impactees responding in desired ways. Evaluators are often expected to provide decision makers with insight into how the intended impactees will react to an intervention, taking into account their bounded rationality and volition, and sociopolitical complexity (Bendor, 2015; Pawson, 2013). The primary objective for this set of games is to test different versions of policy solutions, allowing designers and evaluators to observe the change mechanisms in action, and verify, in close to real conditions, different configurations of the choice architecture that nudge impactees to behave in desired ways. The games include the main rules, positions, and resources assigned to the stakeholders of the intervention, as well as key contextual factors that could influence possible outcomes. Game participants are exposed to the intervention, and
evaluators observe their responses. The primary user of the information obtained is typically the public-sector decision maker.

Exemplar 3: Rural transport regulation game

Policy context, purpose, and audience. Experience with the rural transport regulation game demonstrates how we can use games for ex ante evaluation of proposed complex laws and regulations to inform decision makers. In this case, the Polish national government was the game principal (Olejniczak et al., 2018). In 2017, the Polish government began drafting a new law to respond to the decline of the Polish rural bus transport market and the decreasing accessibility of residents in rural areas to buses. The proposed law would substantially change the complex and fragile dynamics among key stakeholders: the central government, local municipalities, big transportation companies, small local bus operators, and passengers.

In principle, all stakeholders agreed that the regulatory solution should ease the transition from unregulated competition to competition based on public service contracts awarded by the local authorities to bus companies. However, during public deliberations, it became clear that stakeholders were driven by conflicting motivations and held highly divergent views on how the proposed regulation would work in reality and how it would affect them.

A team of evaluators proposed using a game to address two questions: What behaviors will the proposed law trigger among stakeholders? and will the proposed law improve the accessibility of residents in remote rural areas through redevelopment of bus transport? Collecting data to address these two questions was deemed to be helpful for the government’s working group that was preparing an updated draft of the law.

Game content. The game took the form of a 2-day session for 10 players representing different types of stakeholders—local self-government, subregional self-government (powiat), and different bus operators. Passengers’ decisions were simulated since they were assumed to be predictable based on available data and predictive modeling. The game was played in a special decision room with a map representing the region and big boards visualizing the tenders. All participants took roles in the game that were consistent with their positions in real life.

Players handled the transportation system in a subregion consisting of four local communities. Representatives of self-governments designed tenders for transport services taking into account a number of factors. Their objectives were to help children get to school, provide access to buses for residents in isolated villages, and maximize political support for the proposed law. Bus companies analyzed and responded to tenders, competed with different offers, and then delivered services to local governments and competed for commercial customers. The bus operators’ ultimate goal was to maximize profits. Each player could influence the decisions of others and the whole system.

During the first day, participants played according to rules that mirrored the existing law (status quo day). The second day players were introduced to the new rules (proposed law), and they played two rounds that were equal to 2 years in real life. During the entire session, evaluators collected records of all decisions made by all players, including photos of the changing map of the region.

Outcomes of the game. The simulation allowed evaluators to observe the transition from the old to the new law, as well as adaptations and effects, including unintended effects. Four main conclusions from the game were drawn. First, the simulation showed the positive impact of the new regulation on the transport network and accessibility of residents in rural areas, mainly due to the integration of schools and public services. Second, the game showed the positive effect of the proposed law to be highly dependent on the organization and coordination of the tenders. If coordinating mechanisms were not applied by stakeholders early in the process, the whole system could easily get “derailed.” Third, the game revealed the spectrum of bus operators’
strategies, which could be applied or avoided during the transition period. Last, the game results redirected attention from imagined to real obstacles. Attention had been focused on protecting local bus companies because of the fear of bankruptcy, but the game showed instead a need for protecting local governments against price collusions by the operators and defending all stakeholders against a lack of transport during the transitory period if the contracts could not be effectively finalized. Findings of the simulation were used to inform the government committees and policy actors, the game principals in this case, who then used them to inform the design of the draft regulation.

Adaptability. The case of the rural transport regulation game illustrates the possible adaptability of games to meet different objectives. Initially, the game was designed as a testing method of the proposed law. After observing the first day, evaluators added a debriefing at the end of the second day in order to collect more unstructured feedback from the players. It turned out that the debriefing component changed the game into more of a system exploration. Experienced practitioners offered in-depth reflections when asked to talk about their goals, crucial moments in the game, and key factors that affected their decisions. The debriefings expanded all participants’ and observers’ understanding of views on the policy problem from the various stakeholders, helped them to develop more win-win strategies, and offered ideas for improvements in the proposed law.

Exploring Systems

The development of new public policies and programs typically introduces interventions into ecosystems already characterized by complexity. Evaluators are expected to help stakeholders understand the complex system dynamics of the particular problems to be addressed, as well as to manage differences among stakeholders’ mental models and assumptions about both the problems and promising solutions.

This last set of games focuses on collaborative problem-solving. The games can help evaluators and stakeholders engage in dialogue, explore each other’s assumptions about the problem to be addressed, anticipate possible interactions and outcomes that could be triggered by the proposed intervention, and explore possible response scenarios. In these games, evaluators play the role of facilitators of the problem-solving processes in which stakeholders cocreate solutions by testing, exploring, and challenging the systems. Participant learning is the primary goal, not informing game principals.

Exemplar 4: Transforming the Karolinska University Hospital

Policy context, purpose, and audience. Plans were being made to move the Karolinska University Hospital in Stockholm (Solna) into a new hospital building and change its role concept from a general high-quality all-purpose university hospital into a more specialized high-care hospital that would function in a network of other, more general hospitals. The policy was decided upon, the construction procured, and teams were working on the move and design of the new organization. During the planning process, several stakeholders identified the need to explore the system of collaboration around patient logistics, as a key contextual condition for the new hospital to succeed. The set of participants for a game was one of the working groups for the production model of the new hospital, consisting of more than 30 senior health-care professionals from the existing organization. The question to explore was: To what extent are the expectations about how patients’ logistics will work based on valid assumptions about coordination?
Game content. A game was designed as a role-playing game with a primary focus on participants’ learning. Participants were assigned fictional, though realistic roles such as ambulance coordinator and patient coordinator for Hospital X. The assignment to all players was to safely transport all patients in a timely fashion, given existing assumptions about coordination mechanisms, prioritizations, and resource availability. The patient flow was implemented as a set of linked spreadsheets in Google Docs. Patients were dispatched in rounds, and resources (ambulances, medical crew) had to be matched to the patient flow.

There were multiple rounds of play with briefings interspersed throughout the rounds. Playtime was 35 min for a first session and 25 min for a second session (with the same participants in the same roles). The participants were briefed before the session, between the two sessions, and debriefed after the sessions. Given the number of participants, two parallel groups were created, providing the opportunity to compare across the parallel sessions. Between the two sessions, the participants had the opportunity to formulate interventions to improve the system and to reevaluate results of the changes in the second session.

Outcomes of the game. The game immediately revealed the need to further explore the planned capacity (number of ambulances), the coordination mechanisms (when to dispatch patients and when to physically move them in the coordination process), and the vulnerability of the networked hospital system if these items were not further developed before the move into the new Karolinska Hospital. Since the players were from an existing working group for the change process, they were able to use their observations from the game to inform deliberations. The learning from the game informed decision-making, and the move into the new hospital went as planned.

Addressing Challenges of Evaluation With Games

We now discuss in more detail how games may be employed to address three important and persisting challenges in evaluation practice, that is, inclusion, understanding, and utilization. For purpose of emphasis, we separate the three challenges to align aspects of games that are especially relevant to each. However, in practice, characteristics of games, for example, experiential, risk free, providing immediate feedback, are relevant to many of the challenges evaluators face.

The Inclusion Challenge

Inclusion of stakeholders, and especially intended beneficiaries of policies and programs, is an espoused value of evaluation professionals and professions (e.g., see the Guiding Principles of the American Evaluation Association, 2018). But what do we mean by inclusion? Who do we want to include for what? Why does it matter? And why is it so hard to accomplish?

Evaluators typically assume inclusion entails engaging in authentic two-way dialogue with all stakeholders of evaluation, including those least privileged and least likely to have a place at the policymaking table. We accept that it is simply ethically appropriate to respect and listen to all of those involved in evaluation, as noted in the American Evaluation Association (2018) Guiding Principles that state that:

“Evaluators honor the dignity, well-being, and self-worth of individuals and acknowledge the influence of culture within and across groups”, and “[Evaluators] strive to gain an understanding of, and treat fairly, the range of perspectives and interests that individuals and groups bring to the evaluation, including those that are not usually included or are oppositional.”

Inclusion entails both recognizing and respecting differences among stakeholders, but that is not always easy to accomplish effectively. Michael Quinn Patton (2018, p. 22) recently acknowledged
that cultural responsiveness and cultural competence are among those things that evaluators know we should do, but that we are still learning to do well.

A number of evaluation thought leaders, such as Stafford Hood et al. (2015), have been promoting culturally responsive evaluation for years, and they have provided useful methods of inquiry to help evaluators increase their ability to conduct culturally valid work. However, as Ernest House (2017) recently noted, racial framing has significant effects on programs, policies, and evaluation in the United States, and we still are not sufficiently sensitive to the role of racism and bias on the evaluands we confront in our work (p. 167). Both overt bias and unconscious—but ever present—bias affect all stakeholders in evaluation practice—and it is not always easy to recognize its impact (Banaji & Greenwald, 2013).

Inadequate inclusion of important stakeholders—for any reason—can hurt both evaluation processes and results and ultimately limit the validity and usefulness of findings and recommendations from evaluation work. The American Evaluation Association sponsored a series of Dialogues on the Role of Race and Class in our professional work in 2017 and produced an important video that informs evaluation practices but also highlights how daunting the challenge of true inclusion remains (https://www.eval.org/page/racedialogues).

**Benefits of games for inclusion.** Games are, by their very nature, inclusive and equalitarian in the participation of the players. Game masters can be strategic in inviting diverse participants who represent all sorts of diversity and hold differing values relevant to the public policy arena of focus. While inclusion of a diverse set of potential impactees of a policy or program is possible in all types of games, it is perhaps most fruitful in games whose purpose is crash-testing mechanisms or exploring systems so as to ensure that the perspectives of all impactees, not only those with the power and influence, inform policy or program design deliberations. Furthermore, games that explore systems provide a space for stakeholders to articulate their values and rationales for their actions, and such revelations can potentially help prevent or address misunderstandings.

Games are dynamic and reflect the realities that participants face in their work in real time. System exploration games permit revealing the consequences of players’ actions and their system dynamics. The safe game environment provides space for learning how changes in rules and boundaries of the system can include or exclude certain types of players and can be a powerful educational experience to raise awareness and cultural sensitivity. In the Transportation game developed for use in Poland, during the discussion, stakeholders were exposed to explanations of the strategies of other players that improved the public consultation process.

Finally, games can assign players to roles that are different from their actual socioeconomic positions to allow them to look at the world through the eyes of “others,” experience others’ struggles, and spot biases and hidden mechanisms that may lead to harm and injustice. Participants can challenge mental models and values of other participants in a safe space. And participants’ experiences can inform established and proposed solutions and new strategies and models to strengthen inclusion and cultural sensitivity among both evaluators and participants.

**Planning games to address inclusion.** Addressing the inclusion challenge with games requires intentional planning. First, evaluators need to think about inclusion early in the process of game development. The inclusion of pertinent messages can be intentionally integrated early in a game design. For example, noncompetitive versus cooperative game mechanics can expose players to substantially different approaches to thinking about the distribution of resources. Second, evaluators need to bring to the game table all the stakeholders, especially those who may be typically excluded. Again, this aspect has to be considered during game design. Lastly, evaluators need to provide safe space for players’ confrontation of mental models, values sharing, emotionally safe reflection, and exchange of views about what players experience during the game. The discussions and debriefing sessions are
crucial and need to be facilitated well. Those carefully facilitated parts of the session should eventually lead to players’ “a-ha” moments, when they shift their perceptions, and may appreciate the importance of inclusion. Inadequate anticipation of power dynamics among participants or what is needed to create a truly “safe space” for all may backfire, and participants may not feel comfortable expressing themselves candidly.

**The Understanding Challenge**

Public interventions (projects, programs, policies, and regulations) are levers that are designed to activate certain change mechanisms, that in turn should lead to desired effects, and presumably positive change (Chen, 2004; Rossi et al., 1999). However, designers of policy interventions often ignore the existence of and relationships among mechanisms, assuming direct, automatic links between policy actions and the intended beneficiaries’ reactions, and may well neglect side effects. This so-called black box approach to policy design has been widely criticized (Astbury & Leeuw, 2010; Pawson, 2013).

Furthermore, even when unpacking the black box of causal mechanisms, policy practitioners often employ a rational theory of behavior, assuming full rationality of both implementers and intended beneficiaries, that is, ignoring their biases such as over optimism, myopia, illusion of full control, and their unchanging set of preferences. This assumption stands in contrast with the empirical findings of cognitive psychology that reveal how bounded rationality often leads to systematic errors in decision-making (Kahneman, 2011; Simon, 1997; Sunstein, 2000).

Lastly, evaluators continually struggle to more accurately measure policy and program effects. Ex post evaluation often relies on statements of people exposed to the program actions who report about changes in their behavior. More accurate measurement of actual, observed change in behaviors is typically challenging. As David Fetterman pointed out, there is a gap between “theories of action, what we espouse, what we say we’re about, and theories of use. What we actually do, the observed behavior. We need to mind the gap, really close the gap by aligning these things, so we can walk our talk” (Donaldson et al., 2010, p. 40). Therefore, in order to design effective policies and programs, it is crucial to obtain more realistic insights into the change mechanisms that elicit the desired responses from intended beneficiaries (citizens and organizations) to the implemented policy measures (Shafir, 2013; Weaver, 2015). Pretesting to reveal needed adaptations is especially important in an era when politicians, and other observers, are calling for public agents to replicate interventions that have been found to be effective in certain locations elsewhere (Cartwright, 2013).

**Benefits of games for understanding.** Games can reveal how intended and unintended behavioral responses to policies and program mechanisms play out. Similar to the benefits noted for inclusion of stakeholders, and especially impactees who are typically not sufficiently consulted during policy or program design, games create opportunities for evaluators to learn how quite diverse sets of stakeholders will react to a proposed intervention. Testing retention games allow evaluators to assess real, or close to real, behaviors of participants and expose their gaps in skills in a safe environment that doesn’t compromise the stability of system operations. Our examples indicate that games for teaching and testing retention can assist in anticipating implementation challenges and preparing organizations or teams to build their capacity in very concrete skills, for example, knowledge brokering and train traffic control.

Valuable insights can be uncovered through crash-testing and exploratory simulations, as the impactees are the best source of intelligence regarding the mechanics of change in the behavior of individuals and systems. What is more, those two sets of applications provide opportunities for evaluators to observe feedback loops, side effects, and unexpected interactions among actors of the system, its resources, and changing rules. In addition, both crash-test and system exploration games
provide opportunities to speed up time and observe longer term consequences of a policy or program. Both examples discussed above illustrated how game sessions worked as “time machines,” showing players longer term consequences of their decisions and allowing participants to assess sustainability in their expectations of policies. Additionally, system exploration games encourage articulation of hidden, often unconscious, assumptions and allow participants to confront others’ assumptions and values that underpin the messy operations of a particular policy or program in real time. However, it bears noting that these two types of games, contrary to assessing retention games, do not provide precise measurement of effects, but rather indication of potential successes or potential systemic problems.

Planning games for understanding. Our examples demonstrate that effective use of serious games to tackle the understanding challenge requires careful and intentional game design. In the game literature, discussion is focused on three types of validity (Peters et al., 1998): structural validity—degree of isomorphism between the structure of the game and reality, psychological validity—degree to which players perceive the game as realistic and receptive to real-life behaviors, and process validity—degree of isomorphism in actions, decisions, and interactions between the game and reference system. To put it simply, in order to provide accurate findings about explored mechanisms of change, games need to capture key elements of reality, ensure players see the game as realistic, and allow players to behave naturally during the game. To those three aspects of validity, we would add a fourth criteria well known to all evaluators—construct validity. Especially in the case of games for testing retention, it is important to make sure that behaviors we measure in a game environment really represent the desired skills or competencies.

Addressing the validity challenges may be difficult because game designers constantly balance the complexity of reality with limited time (games can’t be too long) and human capacity (rules should be relatively easy to grasp). The examples offered here provide some interesting strategies to deal with these issues. They include: calibrate the appropriate number of sessions or rounds in the game through first conducting pilot sessions run with representatives of the actors of the system being studied, create benchmarks between rounds corresponding to the existing status quo and future scenarios after the intervention, and put time pressures on players so they behave in automatic, intuitive ways which mirror their real-life routines.

Apart from assuring validity, evaluators using games to explore causal explanations need to bring to the table and engage in the game all stakeholders important for the system under inquiry. In the public sector, the use of innovative methods such as simulations may face reluctance. The idea of playing games to evaluate the safety of a multimillion public system (Dutch railways) or assess the rationality of national regulation can indeed sound bizarre to public officials. There are two strategies that may help overcome such pushback. First, use careful wording in the invitation to the game, for example, do not use the term “game,” but rather use a term such as experimental simulation. Second, it is important to clarify to potential participants the benefits that may result from their participation in the exercise. For example, it is helpful to spell out clearly the value of exploring realistic consequences of proposed laws or policies through authentic engagement of a diverse set of stakeholders, especially impactees.

The Utilization Challenge

The raison d’etre of all applied policy research, including evaluation studies, is to be used to inform decision-making processes. However, since the early days of public-sector evaluation work, a disconnect between the production of evaluation studies and use of their findings to inform decision-making has been observed (Weiss, 1988). The body of work on research evidence utilization offers three themes relevant to the use of games to promote learning.
First is the issue of timing. A majority of evaluation studies are undertaken to inform judgments made by decision makers, but the findings, that is, evidence, may well be provided too late in the policy cycle. The findings of ex post evaluations may well arrive after a somewhat revised version of a program or policy has been implemented. And even ex ante evaluation that is designed to predict policy or program outcomes often arrives late in the decision-making process—when decisions have already been made and consensus around the program’s value has been developed. Thus, evaluators frequently face problems in changing the status quo of programmatic operations.

Second, there are important differences between two communities—the producers of knowledge (evaluators, researchers) and the potential users of knowledge (politicians, civil servants, and program managers). The crucial discrepancies relate to different objectives, mind-sets, and languages. Evaluators focus on the production of valid knowledge and finding optimal solutions. Decision makers focus on finding solutions that may feasibly ameliorate pressing problems. Researchers and evaluators tend to explore details and acknowledge uncertainty, while decision makers try to decrease complexity and the uncertainty they cope with every day (Nichols, 2017). And evaluators and researchers tend to use academic language, focusing the narrative on theories and methodological specifics, while practitioners use the language of action and practical implications (Caplan, 1979; Wehrens et al., 2011).

Third, promoting organizational learning is challenging, especially within the context of public-sector decision-making. Individuals and organizational policy actors absorb information and learn in complex, nonlinear ways while established routines tend to promote risk avoidance (Argyris, 1977; Leeuw et al., 1994; Lipshitz et al., 2007). Challenging the status quo of a particular policy requires a body of evidence aligned with other organizational and political processes that eventually could create a situation of punctuated equilibrium for change (Dunn, 2017). Organizational learning tends to be highly incremental, and as Carol Weiss (1980) pointed out—knowledge in the public policy does not flow, it creeps.

**Benefits of games for utilization.** All four of the types of games address communication barriers between the community of research and practice. The universal language of games is a language of practice and action (so-called learning by doing experience). All players in games are transformed into problem solvers and that style is aligned with policy practitioners and more broadly with how adult professionals learn and change their understanding of policy issues and strategies. Furthermore, games are designed as emotionally engaging experiences that offer immediate feedback and that is highly compatible with human learning.

Three areas of game applications—testing retention, crash-testing mechanisms, and exploring systems—clearly promote better timing for learning. The ProRail, transportation regulation, and hospital planning examples show that findings from games preceded actual decisions and investments. The simulations were instrumental in allowing experimentation with different options and anticipating longer term results before things actually happened in reality. Due to the learning through the games, practitioners could tweak the draft design of the policy solutions and test solutions through further games.

The level of engagement of participants in games also has systemic and organizational implications that encourage the incremental process of organizational learning. Evaluation theorists have long advocated that engagement of stakeholders increases the likelihood of utilizations of evaluation results (Shadish et al., 1991), and games present valuable methods for engaging stakeholders for both learning and use. Engaging policy designers and implementers, especially alongside impactees, can enhance their learning about both the programs and policies and the problems they are designed to address.

The engagement offered by games can be instrumental for building three different aspects of shared understanding. First, there is a realization shared among the stakeholders about the necessity
Table 1. Benefits of Games in Addressing the Inclusion, Understanding, and Utilization Challenges in Evaluation Practice.

| Challenge for Evaluation                                                                 | Benefits of Using Serious Games in Addressing the Challenge                                                                 | Things to Consider When Planning Game Application                                                                 |
|-----------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| Inclusion: Involving different actors and perspectives into change process                | (1) Include past, current, or potential impactees of public actions, practitioners, as well as diverse participants and excluded voices who are not typically at the table to provide their input | Consider how the planned game design could reflect and promote the inclusion issue. For example, reflect on choosing cooperative rather than competitive mechanics |
|                                                                                         | (2) Provide safe space for candid dialogue and exchange of experiences and perspectives basing the discussion among participants onto tangible, concrete, yet safe experience—the game session | Construct the list of stakeholders that should be included in the game narrative and/or invited to game sessions. Who are the typical stakeholders? What voices and groups are usually excluded? What are the impactees of the program in a short-term vs. longer term perspective? |
|                                                                                         | (3) Allow participants to assume roles different than their own to experience others’ struggles and perspectives, spot biases, hidden assumptions, and various rationalities | Carefully design debriefing session on inclusion, allowing for emotionally safe discussion on what players experienced during the game. How might participants explain different perspectives and reasons for their actions? How might power dynamics be facilitated in the room? What questions could lead participants towards their “a-ha” moments in challenging their initial assumptions? How might participants be prompted to articulate their key takeaways? What are key ideas on how to translate this experience into policy reality? |
|                                                                                         | (4) Reveal inequitable consequences of policies/programs on different stakeholders and impactees, including implications of how rules and boundaries impact inclusion/exclusion |                                                                                                                  |
|                                                                                         | (5) Build participants’ buy-in for a program or policy changes based on collective social experience |                                                                                                                  |
| Especially promising game types:                                                        | • Teaching knowables                                                                                                       |                                                                                                                  |
|                                                                                         | • Crash-testing mechanisms                                                                                                  |                                                                                                                  |
|                                                                                         | • Exploring systems                                                                                                        |                                                                                                                  |
|                                                                                         | (continued)                                                                                                               |                                                                                                                  |
of change (players see what could happen if the status quo continues). Second, there is shared understanding of the nature of the policy problem (framing the policy issue, identifying key triggers and barriers). Third, there is comprehension of how changes could progress step-by-step (preparing

Table 1. (continued)

| Challenge for Evaluation | Benefits of Using Serious Games in Addressing the Challenge | Things to Consider When Planning Game Application |
|--------------------------|-----------------------------------------------------------|--------------------------------------------------|
| (4) Indicate potential implementation challenges, unintended consequences of new/revised policies, and have a constructive conversation about ways of minimizing those risks | Ensure participation of key stakeholders in a game session and overcome their initial mistrust toward games as serious policy research method. Use terms such as “simulation” or “interactive testing” or “strategic exercise” instead of “game” and spell out the value of exploring realistic consequences of new interventions | |
| (5) Explore systems-level implications of rules and behavioral responses of practitioners and/or impactees | Carefully design debriefing sessions to clarify mechanisms to participants. For testing retention games, debrief participants on their experience. For other games, consider: What questions could prompt participants to articulate their key takeaways and key ideas on improving policy design based on their game experience? | |
| Especially promising game types: | | |
| • Testing retention | | |
| • Crash-testing mechanisms | | |
| • Exploring systems | | |

Utilization: Transferring evaluation insights to policy practice

(1) Promote individual and/or organizational learning through experiential, problem-based, risk-free, and emotionally engaging activities.

(2) Provide safe space for reflection, based on the firsthand experience, hidden assumptions, potential challenges, and risks of new/revised policies and ways of addressing them.

(3) Offer real-time feedback to practitioners on proposed decisions or policies, allow them to anticipate and experience a change in a safe environment before the actual change will be triggered.

(4) Build participants’ shared understanding of the necessity for change, the nature of a policy problem and ways of making a change, and buy-in for a program or policy changes through engagement.

Especially promising game types:

• Teaching knowables
• Testing retention
• Crash-testing mechanisms
• Exploring systems

Plan carefully the timing of the game session to synchronize it with the decision-making cycle of the evaluated intervention. What key decision points do you want to support with the game?

Carefully design game sessions to align with the game purpose and type. What is the representative pool of impactees of the studies interventions? Who are the crucial agents of change in the policy initiative under investigation?

Carefully attend to the narrative and aesthetics of the game design. How can you make the game abstract enough to allow participants to think outside of the box but still resonate with the policy issue in question?

Carefully design debriefing sessions to strengthen learning. Which in-game experiences merit discussion to help build a shared understanding among the participants? How should the discussions be framed to provide participants with practical takeaways?
participants for change in the safe environment). Thus, all four types of games we describe here can be valuable in building a coalition supporting change, enhancing their buy-in when the vetted interventions go live, and securing a smoother transition to new circumstances.

**Planning games for utilization.** Addressing effectively the utilization challenge with games also requires careful planning. First, evaluators need to plan the timing of the game to be in sync with the policy cycle and related decision-making processes. Most of the examples offered here were used as ex ante exercises, early in the policy process, when the logic of the solution was developed but options for change and tweaks were possible. Second, evaluators need to select the participants for the game intentionally. When using games to test solutions, evaluators should include agents of change and draw participants from networks that span across affected organizations. And lastly, game designers need to strategically plan how to incorporate the key policy or program elements into the game narrative and consider physical design and esthetics.

On the one hand, games need to be abstract enough so practitioners can be taken out of their own context to think outside the box, spot more general mechanisms, and not feel threatened. On the other hand, games should resonate with the relevant policy makers and be linked with their reality close enough so they can see the connection between the game and their practice. The approaches used in the examples offered here matched well the real-life challenges, but the physical details and narratives were made to be generic (imagined subregions, generic evaluation units, “typical university hospital”). It is important that game masters draw connections to real-life contexts prior to starting the game and during debriefings. One important caution is that if participants do not understand the relevance of the game to the challenges they face in their work, they may not immerse themselves in the simulation sufficiently.

**Conclusion**

The speed of change in the evolution of the social and environmental problems addressed by public action is breathtaking. Fortunately, ingenuity in the development of serious games to engage, instruct, and explore solutions to the complex problems we face has been rapid as well. We recommend the use of serious games by evaluators to address persisting challenges of inclusion, understanding, and utilization and ultimately make evaluation practice more relevant and impactful.

**Summary of Findings—Games Utility for Evaluation**

We proposed a definition and attributes of “serious games” specifically relevant for application in public policy and program evaluation and distinguished this genre of games from related practice such as gamification. We provided a framework that takes into account two crucial aspects of evaluation inquiry—its function and the nature of the evaluand—and distinguished among four different areas of game applications for evaluation practice: teaching knowables, testing retention, crash-testing mechanisms, and exploring systems. We suggested that games, such as those four types described here, are useful to promote learning about and among stakeholders and to collect valuable intelligence about the operations of programs and policies. Table 1 provides a summary of the benefits of the types of games we describe to tackling the inclusion, understanding, and utilization challenges evaluators face. The table also lists key issues to consider when planning an effective game application for each challenge.

For practice-oriented evaluators, we would like to also raise three issues that are important when planning the application of games in an evaluative inquiry. First, when starting a game project, evaluators should clearly state the purpose of the game and the nature of evaluand. Our framework allows initial mapping of the available choices. This statement will form the core of the so-called design brief, (Kapp, 2012), a document that guides the development of the game and assures that the
initial purpose of the game as an evaluative method will not be lost during the creative, often iterative process of game design.

Second, evaluators need to team up with actual game designers. Dynamic advances in modern programming and Information Technology make it challenging for evaluators to quickly develop expertise that allows professional development of a game. However, evaluators still need to invest time into aligning their language with the jargon used by game designers. Typical game design elements could sound familiar since they are grounded in system thinking language, for example, actors, resources, rules, system boundaries, action arenas, and decision situations, trade-offs, and feedback, but some terms could be challenging, for example, spectrum of mechanics, types of heuristics coded in a game (BoardGameGeek, 2019; Elias et al., 2012; Fullerton, 2019).

Third, when planning the timeline of the study, evaluators should reserve time for calibration and testing of the game prototype. No matter what type of game is employed, game designers will always go through a sequence of “make, break, and remake” that allows them to improve the game by actual experience with real participants.

Future Steps and Directions for Research
Looking forward, more research on the use of serious games in evaluation practice is needed. The positive impact of games on learning and knowledge retention of individuals has been empirically well documented in the literature. However, the question about the effectiveness of the modes of gaming remains unanswered. On the one hand, in the era of the Internet, ubiquitous mobile apps, and personal computers, employing digital games seems an obvious choice. On the other hand, we suggest that analog games offer tangible objects and physical, in-person interaction in games which may be more conducive to human learning. Thus, more comparative research is needed to explore the usefulness of digital versus analog games in the public policy context.

Furthermore, we still know relatively little about the longer term organizational and systemic impact of the use of games in public policy. The cases we described tracked only immediate outcomes after deployment of the games in each particular policy situation. Tracking longer term results of learning facilitated through policy games within and across organizations could contribute to our understanding of knowledge utilization and organizational learning.

Finally, the examples that we provide here demonstrated how games were used to explore complex systems and how they drew upon systems thinking principles. Evaluators are currently developing strategies to address complexity in evaluands and complex behavior more effectively, and game design can benefit from advances in thinking about complex systems.

Games present promising methods of inquiry for evaluators to use in enhancing inclusion, understanding, and utilization in their work. Evaluation practice can benefit from the use of games, and evaluative thinking can contribute to the development of serious games to inform public policy innovation. We hope that we have paved the way to create exciting synergies between evaluation practice and the world of game design.

Acknowledgment
The authors would like to thank colleagues from the evaluation and gaming communities attending the American Evaluation Association 2017 Conference, Australasian Evaluation Society 2018 Conference, and International Simulation and Gaming Association 2019 Conference. Their insightful questions and comments helped in developing this work. The authors would also like to thank the editor and the two anonymous reviewers for their most valuable comments that substantially improved the content of this article.
Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Narodowe Centrum Nauki (grant number 2014/13/B/HSS/03610). Open access of this article was financed by the Ministry of Science and Higher Education in Poland under the 2019-2022 program “Regional Initiative of Excellence”, project number 012 / RID / 2018/19.

ORCID iD
Karol Olejniczak https://orcid.org/0000-0002-7079-2440

References
Abt, C. (1987). Serious games. University Press of America. (Original work published 1969)
Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. American Journal of Evaluation, 38, 434–450. http://doi.org/10.1177/1098214017717015
American Evaluation Association. (2018). American Evaluation Association guiding principles for evaluators. https://www.eval.org/p/cm/ld/fid=51
Argyris, C. (1977). Double-loop learning in organizations. Harvard Business Review, 55, 115–125.
Astbury, B., & Leeuw, F. L. (2010). Unpacking black boxes: Mechanisms and theory building in evaluation. American Journal of Evaluation, 31, 363–381. http://doi.org/10.1177/1098214010371972
Australasian Evaluation Society. (2018, September 19–21). Transformations. International Evaluation Conference, Launceston, Australia. http://conference2018.aes.asn.au/
Baker, M. (2017). Learning agendas: The five most important things you need to know. https://usaidlearninglab.org/lab-notes/learning-agendas-five-most-important-things-you-need-know
Banaji, M. R., & Greenwald, A. G. (2013). Blindspot: Hidden biases of good people. Delacorte Press.
Bardach, E. (2006). Policy dynamics. In M. Moran, M. Rein, & R. E. Goodin (Eds.), The Oxford handbook of public policy (pp. 336–366). Oxford University Press.
Beavis, C. (2017). Serious gaming. In K. Peppler (Ed.), The SAGE encyclopedia of out-of-school learning (pp. 697–698). Sage.
Bendor, J. (2015). Bounded rationality. In J. Wright (Ed.), International encyclopedia of the social & behavioral sciences (2nd ed., pp. 773–776). Elsevier.
BoardGameGeek. (2019). Board game mechanics. https://boardgamegeek.com/browse/boardgamemechanic
Boyle, A. E., Hainey, T., Connolly, M. T., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Pereira, J., & Riberio, C. (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. http://dx.doi.org/10.1016/j.compedu.2015.11.003
Caillios, R. (2001). Man, play and games. University of Illinois Press.
Caplan, N. (1979). The two-communities theory and knowledge utilization. American Behavioral Scientist, 22, 459–470.
Cartwright, N. (2013). Knowing what we are talking about: Why evidence doesn’t always travel. Evidence & Policy: A Journal of Research, Debate and Practice, 9, 97–112. http://doi.org/10.1332/174426413X662581
Charsky, D. (2010). From edutainment to serious games: A change in the use of game characteristics. Games and Culture, 5, 177–198. http://doi.org/10.1177/1555412009354727
Chelimsky, E. (2006). The purposes of evaluation in a democratic society. In I. Shaw, J. C. Greene, & M. M. Mark (Eds.), Handbook of evaluation. Policies, programs and practices (pp. 33–55). Sage.
Chen, H. T. (2004). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Sage.

Christensen, T. (2006). Smart policy? In M. Moran, M. Rein, & R. E. Goodin (Eds.), *The Oxford handbook of public policy* (pp. 448–469). Oxford University Press.

Ciftci, S. (2018). Trends of serious games research from 2007 to 2017. A bibliometric analysis. *Journal of Education and Training Studies, 6*, 18–27. http://doi.org/10.1111/jets.v6i2.2840

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

Costikyan, G. (2013). *Uncertainty in games*. The MIT Press.

Dahler-Larsen, P. (2012). *The evaluation society*. Stanford University Press.

DeSmet, A., Van Ryckeghem, D., Compernolle, S., Baranowski, T., Thompson, D., Crombez, G., Poels, K., Van Lippevelde, W., Bastaensens, S., Van Cleemput, K., Vandebosch, H., & De Bourdeaudhuij, I. (2014). A meta-analysis of serious digital games for healthy lifestyle promotion. *Preventive Medicine, 69*, 95–107. http://dx.doi.org/10.1016/j.ypmed.2014.08.026

de Vaus, D. (2006). Overview. In D. de Vaus (Ed.), *Research design* (pp. 1–38). Sage.

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*, 15–57.

Dörner, R., Göbel, S., Effelsberg, W., & Wiemeyer, J. (2016). Introduction. In R. Dörner, S. Göbel, W. Effelsberg, & J. Wiemeyer (Eds.), *Serious games. Foundations, concepts, and practice* (pp. 1–34). Springer.

Duke, R. (2011). Origin and evolution of policy simulation: A personal journey. *Simulation & Gaming, 42*, 342–358. http://doi.org/10.1177/1046878110367570

Duke, R. (2014). *Gaming: The future’s language*. W. Bertelsmann Verlag. (Original work published 1974)

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

Dunn, W. N. (2017). *Public policy analysis: An integrated approach* (6th ed.). Routledge.

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

Entertainment Software Association. (2018). 2018 essential facts about the computer and video game industry. https://www.theesa.com/esa-research/2018-essential-facts-about-the-computer-and-video-game-industry/

European Evaluation Society. (2018, October 1–5). Evaluation for more resilient societies. Rethinking the role of evaluation in turbulent times. Letter from the Conference Programme Coordinators, Thessaloniki, Greece. http://www.czech-in.org/ees18/download/EES18_Programm_Coordinators_Letter2.pdf

Febey, K., & Coyne, M. (2007). Program evaluation: The board game an interactive learning method of inquiry for evaluators. *American Journal of Evaluation, 28*, 91–101. http://doi.org/10.1177/1098214006296567

Fetterman, D., & Wandersman, A. (2005). *Empowerment evaluation principles in practice*. Guilford.

Fullerton, T. (2019). *Game design workshop: A playcentric approach to creating innovative games* (4th ed.). Taylor & Francis, CRC Press.

Giessen, H. (2015). Serious games effects: An overview. *Procedia—Social and Behavioral Sciences, 17*, 2240–2244. http://doi.org/10.1016/j.sbspro.2015.01.881

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: How effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning, 29*, 207–219. http://doi.org/10.1111/j.1365-2729.2012.00489.x

Graafland, M., Schraagen, J. M., & Schijven, M. P. (2012). Systematic review of serious games for medical education and surgical skills training. *British Journal of Surgery, 99*, 1322–1330. http://doi.org/10.1002/bjs.8819
Gray, D., Brown, S., & Macanufo, J. (2010). *Gamestorming: A playbook for innovators, rulebreakers, and changemakers*. O’Reilly Media.

Grogan, P. T., & Meijer, S. A. (2017). Gaming methods in engineering systems research. *Systems Engineering, 20*, 542–552. http://doi.org/10.1002/sys.21409

Hallsworth, M., Egan, M., Rutter, J., & McCrae, J. (2018). *Behavioural Government. Using behavioural science to improve how governments make decisions*. The Behavioural Insights Team.

Harteved, C. (2011). *Triadic game design. Balancing reality, meaning and play*. Springer.

Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation’s influence on attitudes and actions. *American Journal of Evaluation, 24*, 293–314. http://doi.org/10.1177/109821400302400302

Hood, S., Hopson, R., & Kirkhart, K. (2015). Culturally response evaluation. In K. Newcomer, H. Hatry, & J. Wholey (Eds.), *The handbook of practical program evaluation* (4th ed., pp. 281–317). Jossey Bass.

House, E. (2017). Evaluation and the framing of race. *American Journal of Evaluation, 38*, 167–189. http://doi.org/10.1177/1098214017694963

Iten, N., & Petko, D. (2016). Learning with serious games: Is fun playing the game a predictor of learning success? *British Journal of Educational Technology, 47*, 151–163. http://doi.org/10.1111/bjet.12226

Ivory, J. (2018). *Game studies*. In M. Allen (Ed.), *The SAGE encyclopedia of communication research methods* (pp. 603–607). Sage. https://dx.doi.org/10.4135/9781483381411

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. Pfeiffer.

Kortmann, L. J., & Sehic, E. (2010). The railway bridge game—Usability, usefulness, and potential usage for railways management. *Proceedings of ISAGA, 2010.*

Kortmann, L. J., & Sehic, E. (2011). The Railway Bridge Game – usability, usefulness, and potential usage for railways management. In M. Beran (Ed.), *Proceedings of ISAGA 2010: Changing the world through meaningful play* (pp. 119–125). Eastern Washington University.

Lasswell, H. D. (1951). The policy orientation. In D. Lerner & H. Lasswell (Eds.), *The policy sciences. Recent developments in scope and method* (pp. 3–15). Stanford University Press.

Leeuw, F. L., Rist, R. C., & Sonnichsen, R. C. (1994). *Can governments learn?: Comparative perspectives on evaluation & organizational learning*. Transaction Publishers.

Lipshitz, R., Friedman, V. J., & Popper, M. (2007). *Demystifying organizational learning*. Sage.

Mathison, S. (Ed.). (2005). *Encyclopedia of evaluation*. Sage.

Mayer, I., Bekebrede, G., Harteved, C., Warmelink, H., Zhou, Q., van Ruijven, T., Lo, J., Kortmann, R., & Wenzler, I. (2014). The research and evaluation of serious games: Toward a comprehensive methodology. *British Journal of Educational Technology, 45*, 502–527. http://doi.org/10.1111/bjet.12067

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.

Michael, D., & Chen, S. (2006). *Serious games: Games that educate, train, and inform*. Thomson Course Technology PTR.
Newcomer, K. (2017, November 8). From learning to practice: Employing evaluation to advance the public good. Keynote Address at the Annual Conference of the American Evaluation Association, Washington, DC.

Nichols, T. (2017). The death of expertise. Oxford University Press.

Nightingale, D., Fudge, K., & Schupmann, W. (2018). Evidence toolkit: Learning agendas. Evidence-Based Policymaking Collaborative and Urban Institute. https://www.urban.org/sites/default/files/publication/97406/evidence_toolkit_learning_agendas_2.pdf

Nuttley, S. M., Walter, I., & Davies, H. T. O. (2007). Using evidence: How research can inform public services. Policy Press.

Olejniczak, K. (2017). The game of knowledge brokering: A new method for increasing evaluation use. American Journal of Evaluation, 38, 554–576. http://doi.org/10.1177/1098214017716326

Olejniczak, K., Raimondo, E., & Kupiec, T. (2016). Evaluation units as knowledge brokers: Testing and calibrating an innovative framework. Evaluation, 22, 168–189. http://doi.org/10.1177/1356389016638752

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

Ostrom, E., Gardner, R., & Walker, J. (1994). Rules, games, & common-pool resources. The University of Michigan Press.

Patton, M. Q. (2010). Developmental evaluation: Applying complexity concepts to enhance innovation and use. The Guilford Press.

Patton, M. Q. (2018). Principles-focused evaluation. The Guilford Press.

Pawson, R. (2013). The science of evaluation: A realist manifesto. Sage.

Peters, V., Vissers, G., & Heijne, G. (1998). The validity of games. Simulation & Gaming, 29, 20–30. http://doi.org/10.1177/1046878198291003

Pilon, M. (2015, April 11). The secret history of Monopoly: The capitalist board game’s leftwing origins. The Guardian. https://www.theguardian.com/lifeandstyle/2015/apr/11/secret-history-monopoly-capitalist-game-leftwing-origins

Raghothama, J., & Meijer, S. A. (2018). Rigor in gaming for design: Conditions for transfer between game and reality. Journal Simulation & Gaming, 49, 246–262.

Ramage, M., & Shipp, K. (2009). Systems thinkers. Springer.

Rebolledo-Mendez, G., Avramides, K., de Freitas, S., & Memarzia, K. (2009). Societal impact of a serious game on raising public awareness: The case of FloodSim. Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games, 15–22.

Ricciardi, F., & De Paolis, L. T. (2014). A comprehensive review of serious games in health professions. International Journal of Computer Games Technology, 2014, 1–11. http://dx.doi.org/10.1155/2014/787968

Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). Evaluation: A systematic approach. Sage.

Routledge, H. (2015). Why games are good for business: How to leverage the power of serious games, gamification and simulations. Palgrave Macmillan.

Salganik, M. J. (2018). Bit by bit. Princeton University Press.

Sawyer, B., & Rejeski, D. (2002). Serious games: Improving public policy through game-based learning and simulation. Woodrow Wilson International Center for Scholars.

Sawyer, B., & Smith, P. (2008, February). Serious games taxonomy. Paper presented at the Serious Game Summit, San Francisco, CA.

Schrier, K. (2016). Knowledge games: How playing games can solve problems, create insight, and make change. Johns Hopkins University Press.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). Foundations of program evaluation: Theories of practice. Sage.

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

Simon, H. A. (1997). Administrative behavior (4th ed.). The Free Press.
Snowden, D., & Boone, M. (2007). A leader’s framework for decision making. *Harvard Business Review, November*, 68–76.

Sunstein, C. R. (Ed.). (2000). *Behavioral law and economics*. Cambridge University Press.

Verhagen, H., Johansson, M., & Jager, W. (2017). Games and online research methods. In N. Fielding, R. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 295–306). Sage.

Wang, R., DeMaria, S., Goldberg, A., & Katz, D. (2016). A systematic review of serious games in training health care professionals. *Sim Healthcare, 11*, 41–51. http://doi.org/10.1097/SIH.0000000000000118

Weaver, K. (2015). Getting people to behave: Research lessons for policy makers. *Public Administration Review, 75*, 806–816. http://doi.org/10.1111/puar.12412

Wehrens, R., Bekker, M., & Bal, R. (2011). Coordination of research, policy and practice: A case study of collaboration in the field of public health. *Science and Public Policy, 38*, 755–766.

Weiss, C. H. (1980). Knowledge creep and decision accretion. *Science Communication, 1*, 381–404.

Weiss, C. H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice, 9*, 5–19.

Wenzler, I. (2019, August). *The good, the bad, or somewhere in between: What makes a serious game a seriously good game*. Paper presented at 50th Anniversary Conference of International Simulation and Gaming Association (ISAGA), Warsaw.

Wildavsky, A. (2018). *The art and craft of policy analysis*. Palgrave Macmillan.

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, B., & Hummelbrunner, R. (2010). *Systems concepts in action: A practitioner’s toolkit*. Stanford Business Books.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*, 249–265. http://doi.org/10.1037/a0031311

Young, I. M. (2000). *Inclusion and democracy (Oxford political theory)*. Oxford University Press.

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. http://doi.org/10.1155/2019/4797032