The research and evaluation of serious games: Toward a comprehensive methodology

Igor Mayer, Geertje Bekebrede, Casper Harteveld, Harald Warmelink, Qiqi Zhou, Theo van Ruijven, Julia Lo, Rens Kortmann and Ivo Wenzler

Igor Mayer is an associate professor in Public Administration and Gaming in the Faculty of Technology, Policy and Management, and director of the Serious Gaming Research Center, TU Delft, The Netherlands. Geertje Bekebrede is an assistant professor in Policy, Organization and Gaming in the Faculty of Technology, Policy and Management, TU Delft, and a senior game designer at Tygron Serious gaming, The Netherlands. Casper Harteveld is an assistant professor of Art and Design at Northeastern University. Harald Warmelink is an assistant professor in Policy, Organization and Gaming in the Faculty of Technology, Policy and Management, TU Delft. Qiqi Zhou is an assistant professor in Policy, Organization and Gaming in the Faculty of Technology, Policy and Management, TU Delft. Theo van Ruijven is an assistant professor in Policy, Organization and Gaming in the Faculty of Technology, Policy and Management, TU Delft. Julia Lo is a PhD researcher in Policy, Organization and Gaming in the Faculty of Technology, Policy and Management, TU Delft. Rens Kortmann is an assistant professor in Systems Engineering in the Faculty of Technology, Policy and Management, TU Delft. Ivo Wenzler is an associate professor in Policy, Organization and Gaming in the Faculty of Technology, Policy and Management, TU Delft, and an associate partner at Accenture. Address for correspondence: Dr Igor Mayer, Faculty of Technology, Policy and Management, TU Delft, PO Box 5015, 2600 GA, Delft, The Netherlands. Email: i.s.mayer@tudelft.nl

Abstract

The authors present the methodological background to and underlying research design of an ongoing research project on the scientific evaluation of serious games and/or computer-based simulation games (SGs) for advanced learning. The main research questions are: (1) what are the requirements and design principles for a comprehensive social scientific methodology for the evaluation of SGs?; (2) to what extent do SGs contribute to advanced learning?; (3) what factors contribute to or determine this learning?; and (4) to what extent and under what conditions can SG-based learning be transferred to the real world? In the Netherlands between 2005 and 2012, several hundred SG sessions with 12 SGs were evaluated systematically, uniformly and quantitatively to create a dataset, which comprises data on 2488 respondents in higher education or work organizations. The authors present the research model, the quasi-experimental design and the evaluation instruments. This focus in this paper is on the methodology and dataset, which form a sound foundation for forthcoming publications on the empirical results.

Introduction

The growing interest in (*digital*) *Game-Based Learning* (GBL), *Serious Games* and *Simulation-Gaming* (both abbreviated as SG)[1] entails a growing need to know the effects of what we are doing and promoting. To meet this need, we require proper methods, tools and principles that the fragmented communities agree upon, can validate and apply; in other words, we must take a step

---

[1]The oxymoron Serious Games was first used as title of a book by Clark Abt (1970); the simulation-gaming discipline spurred a.o. by Duke's (1974) *Gaming: the Future's Language*. There's not much light between Abt's and Duke's view on gaming, although later serious games mainly refer to digital nonentertainment games (as artifact). In the context of this research, we will not make a semantic or taxonomic distinction between *serious games* and *simulation games* and will use the abbreviation *SG* for both.

Practitioner Notes

What is already known about this topic

- There are a great many evaluation studies on the efficacy of games for learning.
- There is a plethora of approaches and methods.
- There is an increasing number of reviews on evaluation studies.

What this paper adds

- This publications gives requirements for science and methodology of game-based learning (GBL).
- This paper gives conceptual frameworks and models for comparative evaluation.
- This paper gives structure and examples for comprehensive dataset composition.

Implications for practice and/or policy

- The authors make a strong plea and recommendation to look at evaluation of GBL in a more systematic way.
- The paper gives constructs and items to (re-) use in evaluation of GBL.

---

toward a "science of game-based learning" (Sanchez, Cannon-bowers & Bowers, 2010). It is paradoxical that an increasing number of *older* "DiGRA—Digital Games Research Association," (n.d., "ISAGA," n.d.) and *newly* established institutions in the field ("BGin—Benelux Game Initiative," n.d., "GALA—Network of Excellence for Serious Games," n.d., "SG Academy," n.d., "SGA—Serious Games Association," n.d.) explicitly wish to overcome what they identify as (De Gloria, Bellotti & Berta, 2012):

- geographical fragmentation: aligning for instance countries and geographical markets in the Europe or USA.
- disciplinary fragmentation: aligning disciplines like engineering, humanities and social sciences.
- institutional fragmentation: aligning for instance different research institutes; and
- business chains and networks: aligning for instance producers, consumers, financers, sponsors, publishers, stakeholders, etc.

Considerable efforts and resources are being devoted to researching and evaluating GBL and SGs. As a result, both the number and the quality of such evaluations are increasing (see for a discussion below). However, there are still considerable weaknesses, for example:

- a lack of comprehensive, multipurpose frameworks for comparative, longitudinal evaluation (Blunt, 2006; Meyer, 2010; Mortagy & Boghikian-whitby, 2010; Vartiainen, 2000);
- few theories with which to formulate and test hypotheses (R. E. Mayer, 2005; Noy, Raban & Ravid, 2006);
- few operationalized models to examine "causal" relations (eg, in structural equations models) (Connolly, Stansfield & Hainey, 2009; Hainey & Connolly, 2010);
- few validated questionnaires, constructs or scales, either from other fields (eg, psychology) or newly constructed for SG and GBL (Boyle, Connolly & Hainey, 2011; Brockmyer *et al*, 2009; Mayes & Cotton, 2001);
- a lack of proper research designs, other than randomized controlled trials that can be used in a dynamic, professional learning context (Connolly, Boyle, MacArthur, Hainey & Boyle, 2012; Kato, Cole, Bradlyn & Pollock, 2008; Knight *et al*, 2010; Szturm, Betker, Moussavi, Desai & Goodman, 2011; van der Spek, 2011; van der Spek, Wouters & Van Oostendorp, 2011); and

- an absence of generic tools for unobtrusive ("stealth") data gathering and assessment in and around SGs (Kickmeier-Rust, Steiner & Albert, 2009; Shute, 2011; Shute, Masduki & Donmez, 2010; Shute, Ventura, Bauer & Zapata-Rivera, 2009).

In short, although there is a promising increase in publications, methods, tools and findings, we lack an overarching methodology for SG research. Aligning a fragmented serious gaming community to evaluate and research gaming for learning in a comparative, systematic fashion using procedures, frameworks and methods that can be validated, checked and reproduced represents an enormous challenge.

This is the wider context of our SG research project (2005–12) on the evaluation of SGs for *advanced learning*, which distinguishes it from *Games in K12 education* or *Video games in the classroom*. The challenge is to gather data on the quality, application and outcomes of a broad range of SGs on different topics and with different objectives, used in and for different institutional contexts, at different moments in time and under *uncontrolled* conditions. Although it is valuable to find the effects of playing games with students in a lab, it is essential to know the effects of GBL in uncontrolled circumstances and for objectives that truly matter for real-life performance (ie, emergency management and leadership), as the latter is usually the case in professional learning and training.

In search of a methodology

A social–scientific discipline of SGs research would include a critical and reflective discourse on:

1. frames and discourses: the multiple, often conflicting ways in which we perceive and discuss SGs and GBL (Chong & Druckman, 2007; Shaffer, 2006; Squire, 2002);
2. methodology: the rationale and principles on which SGs and GBL research is founded (Mackenzie & Knipe, 2006);
3. research designs and data gathering: what works, why and when? (De Vaus, 2001; Schneider, 2005);
4. validated research instruments and tools: questionnaires, surveys, logging and tracking instruments, including their validation, for SGs and GBL (Boyle *et al*, 2011; Brockmyer *et al*, 2009; Chertoff, Goldiez & LaViola, 2010; Mayes & Cotton, 2001; Wright & Marsden, 2010);
5. a dynamic body of knowledge identifying the state of the art and knowledge gaps (Ma, Williams, Prejean & Richard, 2007; Raphael, Bachen, Lynn, Baldwin-Philippi & McKee, 2009; Young *et al*, 2012); and
6. professional ethics of the SG designer, the SG advocate, the SG seller, the SG interventionist, etc. (Babbie, 2007; Chandler & Torbert, 2003).

This is highly needed because of:

1. accountability: "users" (clients, players and learners) are becoming more exposed to and familiar with SGs. They have the right to know what they are actually buying, using or playing, for what reason and what the effects or consequences of the application of SGs and *gamification*[2] are. We expect that users will also become more demanding, critical and skeptic; and
2. responsibility: the opposite of accountability. A discipline that advocates the use of SGs and gamification to repair a broken reality (McGonigal, 2011), especially when vulnerable groups in society are involved (eg, children, patients and immigrants) has a great responsibility to critically reflect on the short- and long-term value and structural consequences of the games and gamification tools they are developing, promoting and using.

---

[2]The use of SG can be part of a larger *g*amification approach; but not all forms of gamification include the use of SG. Gamification (McGonigal, 2011) refers to the incorporation of one or more principles of "play" (eg, participant feedback to create competition to give engagement to increase performance, etc.) into (organizational, commercial, social and political) reality.

Research objective

The aim of our project is to develop a generic evaluation methodology for serious gaming that comprises a framework, conceptual models, research design(s), data gathering techniques and evaluation constructs and scales. Our ambition is to resolve the dilemma between the generality and standardization that are necessary for comparative, theory-based research, and the specificity and flexibility that are necessary for evaluating singular cases. The main research questions are:

1. What are the requirements and design principles for a comprehensive methodology for evaluating serious gaming?
2. To what extent do SGs contribute to learning in a real-world context?
3. What factors contribute to or determine this learning?
4. To what extent and under what conditions can serious game-based learning be transferred to the real world?

Cases

Between 2005 and 2012, more than 300 SG sessions with 12 SGs were evaluated as systematically, uniformly, quantitatively and qualitatively as possible, by TU-Delft in cooperation with various partners. All of the games discussed below targeted the learning and training of professionals (to be), in many cases (future) engineers, on things like project management, complex decision making, planning, leadership and team work, combined with specific content and context-related professional knowledge. *Advanced learning* makes declarative, specialized and basic knowledge acquisition less significant than *personal mastery*, combining knowledge, skills and attitudes. Generally, we would not use SG to teach or train university students or professionals to do basic mathematical calculations or what have you, because such things can much better be taught with lectures, books, tutorials and cases. We use SG to let professionals gain more insight into social-technical complexity, and how to handle it for instance when they are put in the position of project leader. This does not make *substance* and *knowledge acquisition* irrelevant because it is the *locus*—not the *focus*—of their professional mastery. It should therefore be included in the narrative and game-play. Many of the games were not only played with university students but also with experienced and senior professionals. All games were played in a "facilitated" mode.

From the research and evaluation efforts of the different games, an evaluation method emerged. Over the years, step by step, we were first reusing evaluation items that worked well. Later, we started testing the items for things like construct validity (scaling, etc.), improved our data collection (see Figure 2 in later section), structured the constructs in an evaluation framework (see Figure 1 in later section), and now leading to structural equation modeling (see, eg, Bekebrede, 2010; Harteveld, 2012; I. S. Mayer, Warmelink & Bekebrede, 2013). At the time of writing, the still expanding dataset contains data on 2488 respondents (comprising male and female students and professionals, aged between 17 and 75) and 960 original variables on such aspects as player background, session, learning and/or policy context, game quality, player experiences, and immediate game results and effects. These data were gathered before, during and after the sessions in a quasi-experimental fashion. Table A1 contains an overview of the background characteristics of the games. Video impressions of four games can be found at http://signaturegames.nl; most other games can be visited on company or game websites. The games are summarized below:

(1) Marine Spatial Planning game (MSP Challenge 2011, TU-Delft, 2011): a computer-supported, multi-player policy game revolving around ecosystem-based, integrated marine spatial planning (MSP) for international professionals working in the field (I. S. Mayer *et al.*, in press; I. S. Mayer *et al*, 2012).
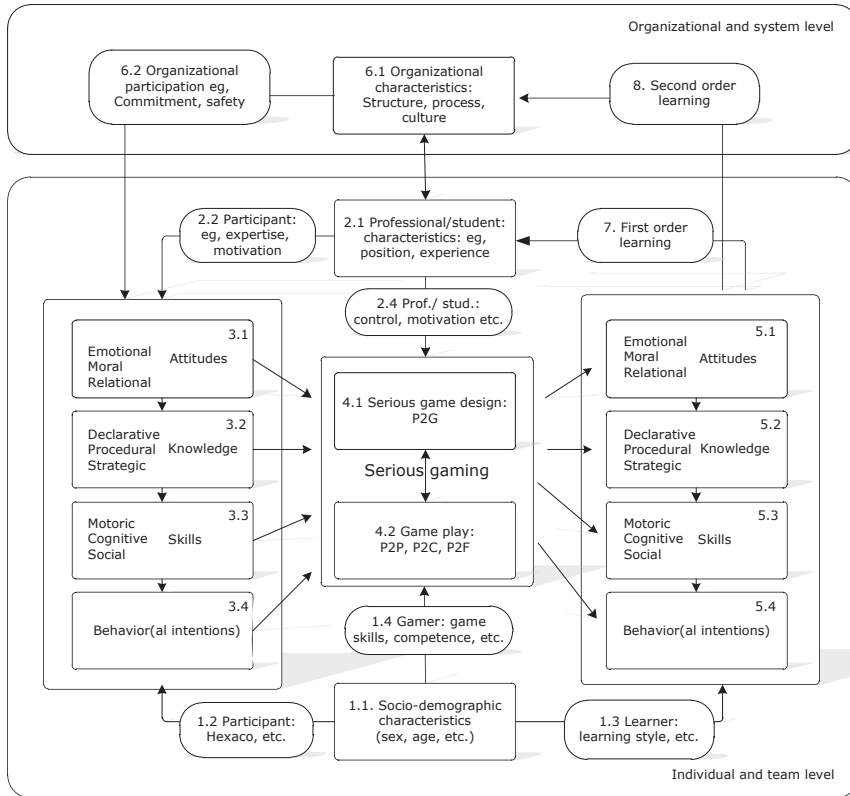
Figure 1: Conceptual framework

(2) SimPort–MV2 (TU Delft, Tygron Serious Gaming, & Port of Rotterdam, n.d.): a computer-supported, multiplayer strategic planning game for higher education and advanced professional learning that revolves around the development of the second Maasvlakte area in the Port of Rotterdam, the Netherlands (Bekebrede, 2010).

(3) Ventum Online ("Ventum Online," n.d.): a computer-supported, multiplayer management game for engineering students and professional project managers that centers on the development of an offshore wind farm.

(4) Construct.it (TU-Delft 2009): a computer-supported, multiplayer planning game for students in higher education concerning the urban reconstruction of a seaport area in a real Dutch town.

(5) Climate Game ("Climategame," n.d.): a 3D computer-supported, multiplayer, strategic planning game for students in higher education and professional policymakers about integrated water management in relation to urban planning, climate change, etc. (Zhou, Mayer, Bekebrede, Warmerdam & Knepflé, 2013).

(6) SprintCity ("SprintCity," n.d.): a computer-supported, multiplayer policy game focused on transit-oriented development around railway stations for professionals working in the field.

(7) Cyberdam (Stichting Rechten Online, 2007): an online, game-based virtual learning environment (VLE) in which teachers in higher education can create their own online role-playing games. Cyberdam is a platform in which 17 different games in 12 institutions have been developed and tested (I. S. Mayer *et al*, 2013).

(8) Servant leadership game (in Dutch: Veerkracht; TU-Delft, 2012): a computer tablet-based, role-playing game about leadership in a changing organization for professionals in a public infrastructure–management organization (Kortmann *et al*, 2012).

(9) Shark World ("Sharkworld," n.d.): a single-player, multimedia, digital game for lower and higher vocational education in project management.

(10) Slogan (Duke, 1981): a nondigital management game for higher education and professional training.

(11) TeamUp (TU-Delft 2010): a digital, 3D, multiplayer game centered on team communication and leadership.

(12) SimVenture (Venture Simulations Ltd, n.d.): a single-player, computer-based game centered on business entrepreneurship.

Table A2 presents the important respondent and background variables in the dataset on the following aspects *per game* and *total for all games*: (1) number of player-respondents (2488 in total); (2) first and most recent dates of play (between 2005 and 2012); (3) nationality (161 non-Dutch); (4) age (between 17 and 75); (5) gender (1650 male; 630 female); (6) student/professional (1831 students, 597 professionals); (7) education (1471 university; 219 middle voc. educ.; 500 higher voc. education); (8) frequency of playing analog games (never–daily); and (9) frequency of playing digital games (never–daily). In addition, and mainly for illustration, we include means (standard deviations) of seven basic multiple item constructs (Cronbach α in Table A2) to give an indication of player satisfaction with the quality of the SG (see Table A2): (1) *clarity* of the SG (five items); (2) *realism* of the SG (two items); (3) *learning satisfaction* (various items); (4) *team engagement* (three items); (5) *facilitator quality* (two items); (6) *user interaction* (two items); and (7) *identification with role*.[3]

Comparative and longitudinal research

In theory, comparative, longitudinal research into GBL has a great many advantages: it provides the opportunity to vary the context or the intermediating variables, such as the number or the intrinsic motivation of students (Paas, Tuovinen, Merriënboer & Aubteen Darabi, 2005), modes of dissemination, the quality or experience of the game facilitators and the institutional facilities (eg, the quality of the classroom or computer infrastructure) (Kenny & Gunter, 2011; Kenny & McDaniel, 2011).

In practice, however, comparative, longitudinal research into GBL presents a great challenge. It requires a high level of discipline and synchronization among the stakeholders, the use of a common evaluation framework that is standardized and robust enough to compare but also flexible enough to allow adjustments to local conditions, institutions, course topics, games and times. It also requires efficient game-based evaluation tools and techniques to gather rich data on a wide spectrum of variables. And while doing the research, the games, the questions, the methods and the tools evolve. In this paper, we discuss the design of the methodology in eight steps, namely:

1. framing;
2. foundations and requirements;
3. conceptual framework;
4. quasi-experimental research design;
5. contextualization;
6. research questions and hypothesis;
7. operationalization; and
8. data reduction and analysis.

[3]For reasons of space and focus, in this paper we do not go into detail about the (validity and reliability) of the items (questions) nor the underlying statistical analysis.

Step 1: framing

Like "learning," the notion "game" in a context of research or science can have different meanings. Consider the differences and possible overlap in the use of "game" as . . .

1. research theory: game theory as in economics, political science, etc. (Leyton-Brown & Shoham, 2008; Shubik, 1999);
2. research concept: organization, management, decision making as a strategic or political game (Scharpf, 1997; Steunenberg, Schmidtchen & Koboldt, 1999);
3. research object: studying game cultures, game economics, game politics, etc. (Castronova, 2005; Ermi & Mäyrä, 2003; Salomon & Soudoplatoff, 2010);
4. design artifact: game as a socio-technical design, as an artifact, etc. (Björk & Holopainen, 2005; Harteveld, 2011; van der Spek, 2011);
5. research method: game as a research method comparable with simulation or experimentation (Barnaud, Promburom, Trebuil & Bousquet, 2007; I. S. Mayer, Carton, de Jong, Leijten & Dammers, 2004; Tykhonov, Jonker & Meijer, 2008);
6. intervention method: game as therapy, educational, learning, change- or decision-support method (Geurts, Duke & Vermeulen, 2007; Preschl, Wagner, Forstmeier & Maercker, 2011); and
7. data gathering method: game as an environment for observation, group interview and data modeling (Cooper *et al*, 2010; Good & Su, 2011; Wood, Griffiths & Eatough, 2004).

Step 2: foundations and requirements

Literature overview

A great many PhD theses and related academic papers on the effects of one or a few GBL and/or serious gaming experiments have now been published (Bekebrede, 2010; Bekebrede, Warmelink & Mayer, 2011; Blunt, 2006; Bremson, 2012; Egenfeldt-Nielsen, 2005; Kato *et al*, 2008; Kuit, 2002; Leemkuil, 2006; Squire, 2004; van der Spek, 2011).

Several review articles on GBL have also been published, and such articles are now appearing with increasing frequency (Adams, 2010; Barlett, Anderson & Swing, 2008; Boyle, Connolly, Hainey & Boyle, 2012; Connolly *et al*, 2012; Coulthard, 2009; Egenfeldt-Nielsen, 2006; Girard, Ecalle & Magnan, 2012; Gosen & Washbush, 2004; Greenblat, 1973; Hays, 2005; Jenson & de Castell, 2010; Ke, 2009; J. Lee, 1999; I. S. Mayer, 2009; Papastergiou, 2009; Randel, Morris, Wetzelf & Whitehill, 1992).

However, few publications provide high-quality evaluation frameworks for what to measure in a comparative fashion and how to do so, taking into account the real-life and dynamic setting of the project (De Freitas & Oliver, 2006). A useful summary and review of 11 evaluation models have been presented by Hainey (Hainey, 2010) and colleagues (Connolly *et al*, 2009; Hainey & Connolly, 2010). Among the 11 frameworks reviewed are the four-dimensional evaluation framework proposed by De Freitas and colleagues (De Freitas & Oliver, 2006; de Freitas, Rebolledo-Mendez, Liarokapis, Magoulas & Poulovassilis, 2010) and Kirkpatrick's four levels for evaluating training (Kirkpatrick, 1994, 2006). This overview is directed at the evaluation of *(digital games) in (formal, K12) education* and aiming at the construction of a model for GBL evaluation (Hainey, 2010). This may explain why other types of evaluation models like the Technology Acceptance Models for serious game adoption (Yusoff, Crowder & Gilbert, 2010) and, more importantly, Kriz and Hense's framework for theory-based evaluation—used for simulation-games, also nondigital—are not taken into consideration (Bekebrede, 2010; Kriz & Hense, 2004, 2006).

Limitations of existing frameworks

The models and frameworks discussed above are high-level models. They specify a limited number of generic concepts that can or should be taken into consideration when evaluating SGs. With relevance in their own right, models like De Freitas' *four dimensional framework* or Kirkpatrick's

four levels of evaluation are not easy to use for exploratory or explanatory hypothesis generation and testing. Moreover, there are even fewer evaluation frameworks of GBL in *higher education*, let alone professional, in-company training or group and organizational learning. In sum, the existing models (we know) give:

1. few indications how to use the models, for what purpose, with what scope and under what conditions;
2. few procedures to validate the conceptual research/evaluation model;
3. few research hypotheses and research designs;
4. few definitions of or relations and interrelations between the concepts in the model; and
5. few operationalizations and validations of constructs.

Furthermore, in the application of the models, we see:

1. a dominance of single case-studies, one game, one context of application;
2. a lack of information on the questionnaires used;
3. a focus on the GBL of children in formal education; little attention to advanced–professional learning outside education; and
4. a focus on the learning of individuals in formal training or the educational context; little attention to the learning of teams, groups, organizations, networks or systems in a policy or organizational context.

Requirements
An important question, therefore, is what the requirements are for a good evaluation framework for serious game evaluation research. A generic evaluation framework (and corresponding procedures), for GBL and SGs research ideally (and based upon our experience with the evaluation of the 12 games between 2005 and 2012) has the following characteristics:

1. broad scope: takes into account the broad range of educational contexts, games, learning objectives and topics;
2. comparative: able to use certain data from different games for comparison;
3. standardized: in order to use a pre-/quasi-experimental research design, materials and procedures should be standardized;
4. specific: measuring data precisely by pinpointing variables;
5. flexible: as game play cannot be always predicted, data gathering should be flexible for measurement; however, still standardized, specific, etc.;
6. triangulated: using a mixed method approach with qualitative and quantitative data;
7. multileveled: individual, game, team, organization and system levels;
8. validated: validated research methods, for example, research method and game design;
9. expandable: possibility of measurements on new variables;
10. unobtrusive: using gaming for systematic and extensive data gathering (research, comparative or theory-based evaluation, etc.) needs to be unobtrusive;
11. fast and nontime consuming: using real-world cases for data gathering implies that tools and methods need to be fast and nontime consuming, because in real-world projects, not much time and resources can be devoted to research; and
12. multipurposed: persuading stakeholders to do data gathering beyond the obvious and minimal.

In practice, no such framework exists and trade-offs need to be made. A GBL or an SG evaluation needs to be broad in scope but light in operation; it must address both the formative and the summative purposes of evaluation (Bloom, Hastings & Madaus, 1972) and the evaluation interests of the designers, players, financers and other stakeholders. At the other end of the spectrum, the data need to be suitable for more in-depth analysis, in order to understand what happens and why.

Step 3: conceptual framework

Elements

A generic model for the social scientific research, evaluation and assessment of SGs in a real-world context should provide:

1. a flexible and generally applicable research model from which we can derive:
    (a) a set of research questions and hypotheses;
    (b) a research design for applying the model;
    (c) a suite of research tools and instruments; and
    (d) some guidelines, practices and rules for applying, falsifying, validating and improving the above.
2. Empirical testing of the robustness of the model.

Comparative, theory-based evaluation

Before the learning effectiveness and contributing factors can be established, an evaluation framework is required that allows:

(1) the operationalization of independent, dependent and mediating or context variables, like "engagement" (in this case, independent), "learning effectiveness" (in this case, dependent) and age (mediating), or psychological safety (context);
(2) a systematic, unobtrusive process of data gathering and data analysis; and
(3) the formulation of research questions and hypotheses based on a conceptual research model.

The core of the model is a deconstruction of GBL into:

1. The pregame condition: the subject's attitudes, knowledge, skills and behavior relevant to GBL and SGs and/or the case at hand before playing the game. In the 12 cases, we measured a variety of items and constructs ranging from, for example, attitudes toward GBL to organizational commitment (see 3.1–3.4 in Figure 1).
2. The quality of the GBL intervention: subdivided into the quality of the game design itself, the game play, the interaction with the facilitator/ teacher and the interaction with the digital game environment (see 4.1–4.2 in Figure 1).
3. The postgame condition: the subject's attitudes, knowledge, skills and behavior relevant to the GBL etc. (see 5.1–5.4 in Figure 1).
4. Background variables referring to the person, student or professional as a participant:
    (a) Socio-demographic variables: age, sex, nationality, etc. (see 1.1 in Figure 1).
    (b) Professional and student characteristics: position, work experience, level of education, etc. (see 2.1 in Figure 1).
5. Mediating variables
    (a) Individual as a participant (eg, personality characteristics; Big 5, Hexaco) (see 1.2 in Figure 1).
    (b) Individual as a learner (learning styles, etc.) (see 1.3 in Figure 1).
    (c) Individual as a gamer: (eg, game skill, game experiences, game attitudes, game-play style, etc.) (see 1.4 in Figure 1).
    (d) Professional/student as a serious gamer (eg, previous experiences with SGs in a professional context) (see 2.4 in Figure 1).
    (e) Professional/student as a participant (eg, intrinsic/extrinsic motivation) (Ainley & Armatas, 2006).
6. Context variables: organizational/institutional climate in which the GBL/SG takes place (eg, commitment to the organization, identification with leader or organization, psychological safety) (see 6.1 in Figure 1).
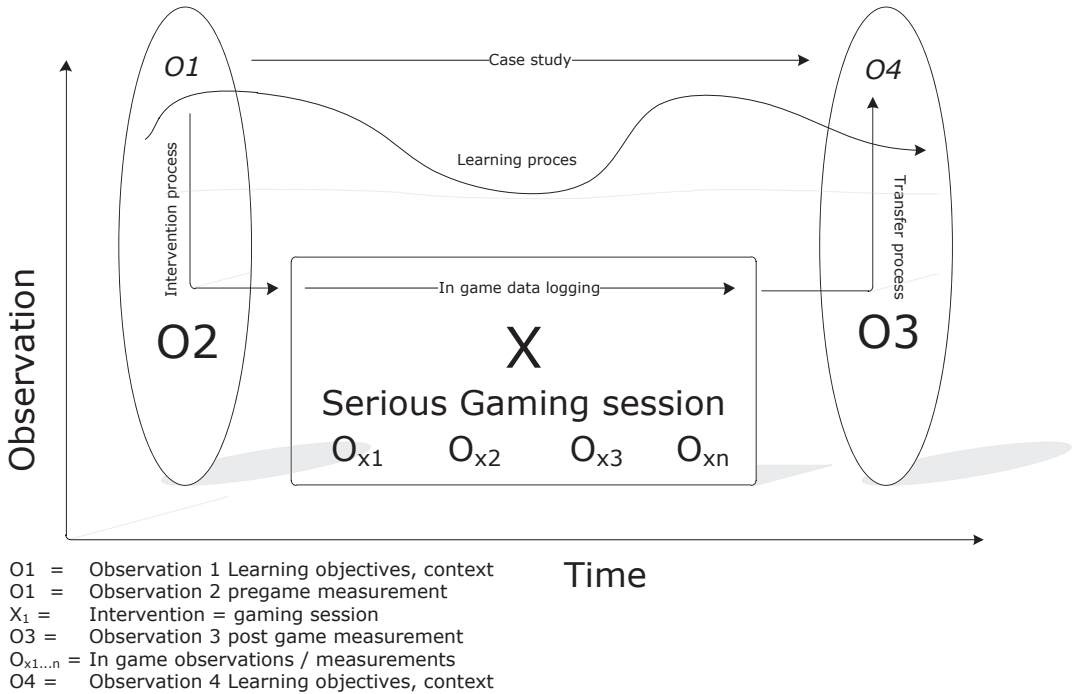
O1 = Observation 1 Learning objectives, context
O1 = Observation 2 pregame measurement
X_1 = Intervention = gaming session
O3 = Observation 3 post game measurement
$O_{x1...n}$ = In game observations / measurements
O4 = Observation 4 Learning objectives, context

Figure 2: Generic quasi-experimental design for GBL and SGs

7. First-order learning: direct influence of playing the game on the individual, small group attitudes, knowledge, skills or behavior (see 7. In Figure 1).
8. Second-order learning: direct/indirect, short- or long-term influence of the game in the large (including design process, sessions, discussions, publications, other interventions, etc.) on the group, network, organizational and system levels (see 8 in Figure 1).

Step 5: quasi-experimental research design
Now the model can be translated into a quasi-experimental design: from the simple "post-test only," to a "pre-test/post-test" design, a "randomized (R)," "control group (C)," "repeated measurement" design (Campbell & Stanley, 1963; Cook & Campbell, 1979; Creswell, 2002). Figure 2 illustrates the basic translation of the conceptual model into a quasi-experimental design (R and C not included in figure):

Step 6: contextualization
Data gathering
One of the special features of SGs for advanced learning is that they provide excellent environments for mixed method data gathering, or triangulation: from crowd sourcing to panel discussions, surveys and observations/video observations. Figure 3 gives a visual impression of methods that can be mixed with SGs.

For the games listed in Table A1, evaluation data were gathered through mixed methods, mostly combining pregame and postgame surveys among the players, live or video observations, transcripts of after-action reviews and game results. In a few cases, methods were applied more rigorously with in-game knowledge tests or network and communication analyses from video
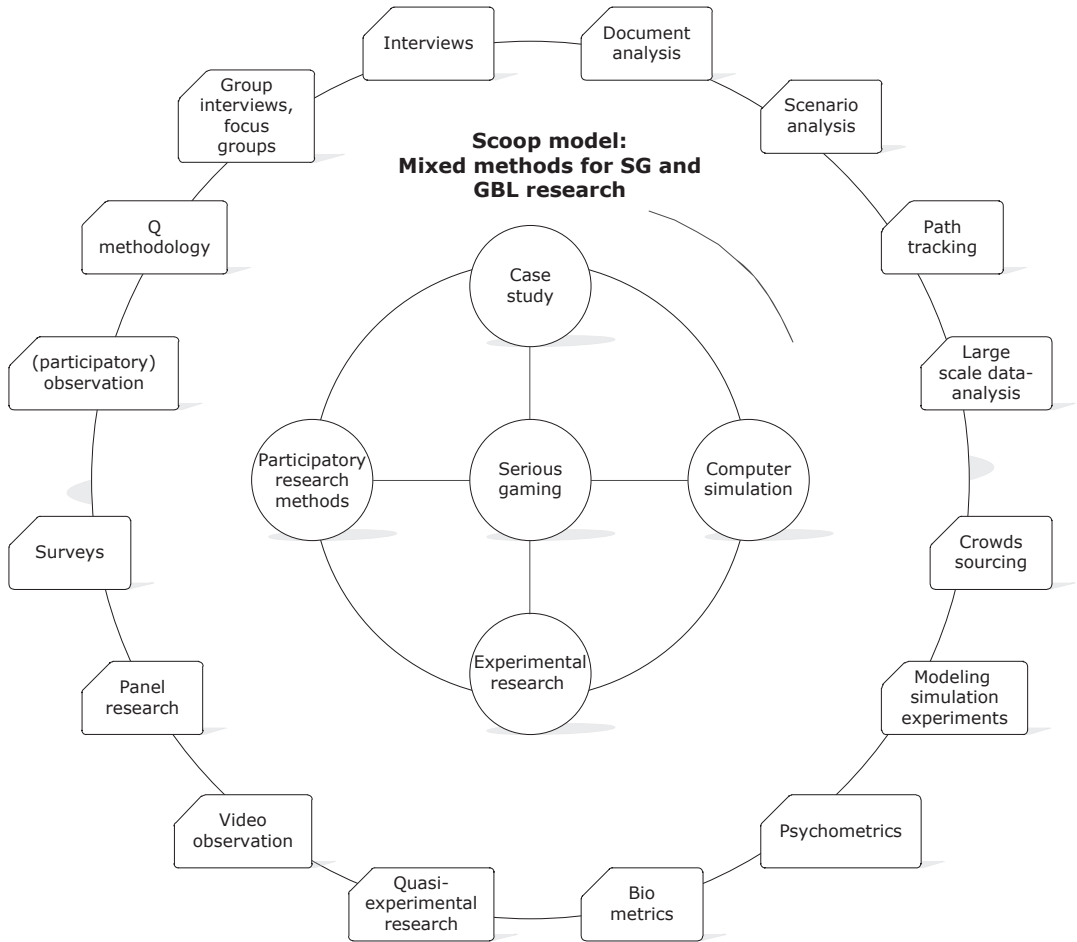
Figure 3: SGs and data gathering methods

observations. Table 1 gives an overview of how to mix the various methods in pregame, in-game and postgame stages.

Step 6: defining questions and hypotheses
We classify the types of research questions and research hypotheses that can guide GBL and SGs research into:

1. design-oriented research (artifact): "making it (better)":
   (a) The validation of specific and generic game-based artifacts and events; and
   (b) The development and validation of design theories, methods and tools.
2. Intervention-oriented research (learning, change, policymaking, management): "making it work":
   (a) The learning effectiveness/impact of game-based interventions; and
   (b) The transfer of game-based interventions to the real world.
3. Domain-oriented research (healthcare, military, energy, etc.): "making it matter":
   (a) The effectiveness of the use of SGs to understand the complexity, dynamics in specific domains.

Table 1: What to measure, how and when

How		What?	Pregame	In-game	Postgame
Self-reported	Qual.	Personality, player experiences, context, etc.	Interviews, focus group, logbook.	Logbook, interviews or small assignments as part of the game.	Interviews focus group, after-action review.
	Quant.	Socio-demographic, opinions, motivations, attitudes, engagement, game-quality learning, power, influence, reputation, network centrality, learning satisfaction, etc.	Survey, quest., individual or expert panel	In-game questionnaires	Survey, quest., individual or expert panel
Tested	Qual.	Behavior, skills, etc.	Eg. actor role-play, case-analysis, assessment, mental models.	Game-based behavioral assessment.	Game-based behavioral assessment.
	Quant.	Values, knowledge, attitudes, skills, personality, power.	Psychometric, socio-metric tests: eg, personality, leadership, team roles, IQ.	Game-based behavioral performance analysis.	Game-based behavioral performance analysis.
Observed	Qual.	Behavioral performance of student, professionals, player and/or facilitator, others; decisions, strategies, policies, emotions, conflicts, etc.	Participatory observation, ethnographic methods	Video, audio personal observation, ethnography, maps, figures, drawings, pictures, etc.	Participatory observation, ethnographic methods.
	Quant.	Biophysical–psychological responses, like stress (heart freq., perspiration).	Part. observation, network analysis, Biophysical–psychological observation.	In-game tracking and logging, network analysis, data mining, biometric observation.	In-game log file analysis, network analysis.

4. Disciplinary research (methodology, ethics, explanatory and interpretative theories): "making it understandable":
   (a) he sociological, economic, political, cultural, etc. frames on SGs;
   (b) Theory construction on GBL and SGs;
   (c) Methodology: design and validation of research methods and tools; and
   (d) Reflection and ethics.

Depending upon the case at hand, a pregame, an in-game and a postgame instrument for measuring or observing relevant variables can now be defined or constructed. There is a great variety of games, players and learning contexts, and trade-offs need to be made between time, resources and the focus of the evaluation (see above). The first-order effects can be established as changes between pregame and postgame measurements, with or without a control group.

Knowledge acquisition for instance can be evaluated with *knowledge tests* (eg, exam style) making it even stronger when combined with *self-reporting* and, or *in-game performance measurements*. An excellent example of such a multimethod assessment of knowledge acquisition among professionals can be found in Harteveld (2012). The ex-post, self-reporting or self-assessment of change and learning however is quite common and often sufficient. Recent studies found a strong and significant correlation between self-reported learning, in-game performance improvement and pregame, postgame testing (Harteveld, 2012). Yet, even when based upon self-reporting, high-quality questionnaires with items, constructs and scales for comparative and longitudinal measurements of knowledge acquisition and learning are not commonly available.

Step 7: operationalization

The operationalization of the generic conceptual model (Figure 1) in the context of a dynamic, multistakeholder project can be a real challenge. First, most educational or client organizations have their own procedures and preferences for evaluations; sometimes a set of course evaluation questions is mandatory. Paying clients are not always inclined to evaluate beyond their own immediate needs (eg, "Did we get what we expected and were promised?"). In many cases, we need to convince stakeholders that for the purpose of advancing serious gaming, a thorough evaluation should be done. We need to emphasize and ensure privacy, safety and nonobtrusiveness. Another handicap is that the client usually expects case-specific evaluation questions. Ad hoc and case-specific evaluation questions stand in the way of comparative, longitudinal research. Through the years, we have found flexible ways of working, trying to validate and reuse as much as possible the items, constructs and scales. In our comparative research (Tables A1 and A2), we have gradually build up a set of validated and reusable questions for the following constructs and items:

Pregame

1. Socio-demographics: sex, age, nationality, culture, etc. (Bekebrede *et al*, 2011; Boyle & Connolly, 2008; D. J. Brown, Ley, Evett & Standen, 2011; M. Brown, Hall, Holtzer, Brown & Brown, 1997; Erfani *et al*, 2010; G. Hofstede, 1986; Jenson & de Castell, 2010; Kinzie & Joseph, 2008; Pfister, 2011).
2. Previous experiences/skills: with computers, games, VLEs, etc. (Erfani *et al*, 2010; Harper *et al*, 2007; Mortagy & Boghikian-whitby, 2010).
3. Attitudes: change, conflicts, intrinsic and extrinsic motivation, learning styles, etc. (Ashton & Lee, 2009; Garris, Ahlers & Driskell, 2002; Guay, Vallerand & Blanchard, 2000; Huang, 2011; K. Lee & Ashton, 2004; Lepper & Malone, 1987; Malone & Lepper, 1987).
4. Skills: personal competence (with games, learning, certain professional skills, etc.) (D. J. Brown *et al*, 2011; Enochsson *et al*, 2004; Holsbrink-Engels, 1998; Verdaasdonk *et al*, 2009; Wolfe & Box, 1988).
5. Behaviors (behavioral intentions)

6. Group, team, organizational characteristics: team/group conflict, psychological safety, psychological collectivism, team and organizational commitment, etc. (Brockner & Higgins, 2001; Carmeli, Brueller & Dutton, 2009; Edmondson, 1999; Ferris, 2005; Jackson, Colquitt, Wesson & Zapata-Phelan, 2006).

In-game

7. Game performance: based upon in-game scores, such as time, avoidable mistakes, etc. (Baba, 1993; Blumberg, 2000; Oslin, Mitchell & Griffin, 1998; Tallir, Lenoir, Valcke & Musch, 2007; Trepte & Reinecke, 2011).
8. Game play: effort; dominance, influence, power, etc.
9. Game experience: flow, immersion, presence, etc. (Admiraal, Huizenga, Akkerman & Dam, 2011; Csikszentmihalyi, 1991; Martin & Jackson, 2008).

Postgame

10. Game experience: engagement, fun while playing the game, etc. (Boyle *et al*, 2012; Mayes & Cotton, 2001; Schuurink, Houtkamp & Toet, 2008).
11. Player satisfaction with: the game (eg, clarity, realism); user interaction (eg, attractiveness, ease of use, computer malfunctions, support, etc.); the quality of the facilitator (eg, supportive, player identification with facilitator); the interaction with other students (eg, player efforts, motivation); identification of players with role; team engagement (Olsen, Procci & Bowers, 2011; Reichlin *et al*, 2011) (see also Table A2 and below).
12. First-order learning (short term, individual, participants).
    (a) Player learning satisfaction, self-reported, self-perceived learning, for example broad range of items.
    (b) Measured changes in knowledge, attitudes, skills and behaviors (behavioral intentions).
13. Second-order learning (medium term, long term, collective, participants and nonparticipants):
    (a) Self-reported, case-based, reconstructive: asking clients, participants, etc. how the results of the GBL have been implemented.
    (b) Measured changes in team, group, organizational characteristics: safety, commitment, performance, performance, etc.

Step 8: data reduction and analysis

A final and important concern is data reduction and analysis. Over the years, we have varied and changed items, questions and constructs. Data reductions through factor analysis and reliability analysis of scales increasingly allow us to select the influential and distinguishing items and construct scales. The present dataset now contains a total number of 960 variables about the 12 games.

As indicated above, we frequently use the same scales or constructs to measure "game design quality" and "player satisfaction" through constructs like clarity, realism, facilitator quality, user interaction etc. Table A2 gives an indication of how the 12 games individually, on average and comparatively score on seven basic constructs. The end goal of comparative research is to test the efficacy of GBL and SGs through *structural equation models* (forthcoming publications; see also I. S. Mayer *et al*, 2013) using also validated or newly constructed psychometric scales and constructs for the broad range of constructs listed above.

Conclusions

We have formulated both the requirements for and a conceptual research model that can be translated into quasi-experimental research designs and operationalized into an evaluation model for specific cases and contexts of GBL. We have also demonstrated the principles and workings of the model on the basis of a comparative case of 12 SGs. The strengths of our model

we believe is that it finds a balance between broad application scope through flexibility and cross-case comparative research through a modular evaluation framework. The framework does not rigidly superimpose hypotheses about sGBL but does support the generation, exploration and testing of hypothesis and structural models (I. S. Mayer *et al*, 2013).

We are aware of at least one weakness in our approach to evaluate the 12 games: we had multiple roles and potential interests in their evaluation as designers, users, teachers, facilitators, entrepreneurs, etc. We have done our best over the years to separate our roles as much as possible and to approach evaluation as systematically and critically as conditions allowed. In our role as designers or teachers, we would like to see confirmed that the games we designed and used were engaging and effective learning tools; as researchers and evaluators however we had to live with the fact that sometimes they were not. For the designer or teacher this can be painful; for the researcher it raises questions about if, what, when and how serious gaming works (G. J. Hofstede, De Caluwe & Peters, 2010). By and large, we see three ways forward in our research:

1. to perform a comparative analysis of the data in order to find the factors that influence the efficacy of GBL and SGs;
2. to enrich and improve the constructs and scales for GBL and SGs; and
3. to use a digital tool for the quasi-experimental design research into SGs and GBL that allows the immediate coupling of a variety of pregame, in-game and postgame data.

Acknowledgements

References

Abt, C. (1970). *Serious games*. New York: Viking Press.

Adams, S. A. (2010). Use of "serious health games" in health care: a review. *Studies in Health Technology and Informatics*, *157*, 160–166. Retrieved May 3, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/20543383

Admiraal, W., Huizenga, J., Akkerman, S. & Dam, G. T. (2011). The concept of flow in collaborative game-based learning. *Computers in Human Behavior*, *27*, 3, 1185–1194. doi:10.1016/j.chb.2010.12.013.

Ainley, M. & Armatas, C. (2006). Motivational perspectives on students ' responses to learning in virtual learning environments. In J. Weiss, J. Nolan, J. Hunsinger & P. Trifonas (Eds), *The international handbook of virtual learning environments* (pp. 365–394). Dordrecht, the Netherlands: Springer. doi: 10.1007/978-1-4020-3803-7_15.

Ashton, M. C. & Lee, K. (2009). An investigation of personality types within the HEXACO personality framework. *Journal of Individual Differences*, *30*, 4, 181–187. doi:10.1027/1614-0001.30.4.181.

Baba, D. M. (1993). Determinants of video game performance. (J. L. Starkes & F. Allard, Eds.) *Advances in Psychology*, *102*, 57–74. Retrieved May 3, 2013, from http://ezproxy.library.unlv.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1993-98813-004&site=ehost-live

Babbie, E. (2007). The ethics and politics of social research. In E. Babbie (Ed.), *The practice of social research* (pp. 60–83). London: Thomson Wadsworth.

Barlett, C. P., Anderson, C. A. & Swing, E. L. (2008). Video game effects—confirmed, suspected, and speculative: a review of the evidence. *Simulation & Gaming*, *40*, 3, 377–403. doi:10.1177/1046878108327539.

Barnaud, C., Promburom, T., Trebuil, G. & Bousquet, F. (2007). An evolving simulation/gaming process to facilitate adaptive watershed management in northern mountainous Thailand. *Simulation & Gaming*, *38*, 3, 398–420. doi:10.1177/1046878107300670.

Bekebrede, G. (2010). August 24). *Experiencing complexity: a gaming approach for understanding infrastructure systems*. TU Delft. Retrieved May 3, 2013, from http://www.narcis.nl/publication/RecordID/oai:tudelft.nl:uuid:dae75f36-4fb6-4a53-8711-8aab42378878

Bekebrede, G., Warmelink, H. J. G. & Mayer, I. S. (2011). Reviewing the need for gaming in education to accommodate the net generation. *Computers & Education*, 57, 2, 1521–1529. doi:10.1016/j.compedu. 2011.02.010.

BGin—Benelux Game Initiative (n.d.). Retrieved December 20, 2012, from http://www.bgin.org/

Björk, S. & Holopainen, J. (2005). *Patterns in game design*. p. 423. Boston: Charles River Media.

Bloom, B. S., Hastings, T. J. & Madaus, G. F. (1972). Handbook on formative and summative evaluation of student learning. *Studies in Art Education*, 14, 68–72.

Blumberg, F. C. (2000). The effects of children's goals for learning on video game performance. *Journal of Applied Developmental Psychology*, 21, 6, 641–653. doi:10.1016/S0193-3973(00)00058-7.

Blunt, R. D. (2006). *A causal-comparative exploration of the relationship between game-based learning and academic achievement: teaching management with video games*. Minneapolis, MN: Walden University. Retrieved May 3, 2013, from http://www.rickblunt.com/phd/blunt_richard_dissertation_final.pdf

Boyle, E. A. & Connolly, T. M. (2008). Games for learning: does gender make a difference? *2nd European Conference on Games Based Learning*, 69–76524.

Boyle, E. A., Connolly, T. M. & Hainey, T. (2011). The role of psychology in understanding the impact of computer games. *Entertainment Computing*, 2, 2, 69–74. doi:10.1016/j.entcom.2010.12.002.

Boyle, E. A., Connolly, T. M., Hainey, T. & Boyle, J. M. (2012). Engagement in digital entertainment games: a systematic review. *Computers in Human Behaviour*, 28, 3, 771–780.

Bremson, J. (2012). Using gaming simulation to explore long range fuel and vehicle transitions, University of California, California. Retrieved May 3, 2013, from http://steps.ucdavis.edu/its_publications/dissertations/Bremson_Joel.pdf

Brockmyer, J. H., Fox, C. M., Curtiss, K., McBroom, E., Burkhart, K. M. & Pidruzny, J. N. (2009). The development of the Game Engagement Questionnaire: a measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, 45, 4, 624–634. doi:10.1016/j.jesp.2009.02.016.

Brockner, J. & Higgins, E. T. (2001). Regulatory focus theory: implications for the study of emotions at work. *Organizational Behavior and Human Decision Processes*, 86, 1, 35–66. doi:10.1006/obhd.2001. 2972.

Brown, D. J., Ley, J., Evett, L. & Standen, P. (2011). Can participating in games based learning improve mathematic skills in students with intellectual disabilities? *2011 IEEE 1st International Conference on Serious Games and Applications for Health SeGAH* (pp. 1–9). IEEE. doi:10.1109/SeGAH.2011.6165461.

Brown, M., Hall, L. R., Holtzer, R., Brown, S. & Brown, N. (1997). Gender and video game performance. *Sex Roles*, 36, 11, 793–812. doi:10.1023/A:1025631307585.

Campbell, D. T. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago: Rand McNally.

Carmeli, A., Brueller, D. & Dutton, J. E. (2009). Learning behaviours in the workplace?: The role of high-quality interpersonal relationships and psychological safety. *Systems Research and Behavioral Science*, 26, November 2008, 81–98. doi:10.1002/sres.

Castronova, E. (2005). *Synthetic worlds: the business and culture of online games*. Chicago: University of Chicago Press.

Chandler, D. & Torbert, B. (2003). Condescending ethics and action research. *Action Research*, 1, 2, 37–47. Retrieved May 3, 2013, from http://arj.sagepub.com/cgi/content/abstract/4/1/37

Chertoff, D. B., Goldiez, B. & LaViola, J. J. (2010). Virtual Experience Test: a virtual environment evaluation questionnaire. *2010 IEEE Virtual Reality Conference VR* (pp. 103–110). IEEE. doi:10.1109/VR. 2010.5444804.

Chong, D. & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, 10, 1, 103–126. doi:10.1146/annurev.polisci.10.072805.103054.

Climategame (n.d.). *(Serious urban planning and water management game)*. The Hague: Tygron. Retrieved May 3, 2013, from http://climategame.nl/

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

Connolly, T. M., Stansfield, M. & Hainey, T. (2009). Towards the development of a games-based learning evaluation framework. *Games-based learning advancements for multisensory human computer interfaces: techniques and effective pratices*.

Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Boston: Houghlin Mifflin Company. Boston, MA: Houghton Mifflin. (p. 405). Retrieved May 3, 2013, from http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0395307902

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M. *et al* (2010). Predicting protein structures with a multiplayer online game. *Nature*, 466, 7307, 756–760. doi:10.1038/nature09304. Retrieved

May 3, 2013, from http://www.nature.com/nature/journal/v466/n7307/abs/nature09304.html#supplementary-information

Coulthard, G. J. (2009). *A review of the educational use and learning effectiveness of simulations and games. Business.* West Lafayette, IN: Purdue University. Retrieved May 3, 2013, from http://www.coulthard.com/library/paper-simulation.html

Creswell, J. W. (2002). *Research design: qualitative, quantitative, and mixed methods approaches (2nd edition). Organizational research methods.* Vol. 6 (p. 246). Thousand Oaks, CA: Sage Publications, Inc. Retrieved May 3, 2013, from http://www.amazon.com/Research-Design-Qualitative-Quantitative-Approaches/dp/0761924426

Csikszentmihalyi, M. (1991). *Flow: the psychology of optimal experience.* New York: Harper Perennial.

De Freitas, S. & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education*, *46*, 3, 249–264. Retrieved May 3, 2013, from http://www.sciencedirect.com/science/article/B6VCJ-4HVDYJY-2/2/821ec9dbad4ba57a1f779f62ba2b0e57

De Freitas, S., Rebolledo-Mendez, G., Liarokapis, F., Magoulas, G. & Poulovassilis, A. (2010). Learning as immersive experiences: using the four-dimensional framework for designing and evaluating immersive learning experiences in a virtual world. *British Journal of Educational Technology*, *41*, 1, 69–85. doi:10.1111/j.1467-8535.2009.01024.x.

De Gloria, A., Bellotti, F. & Berta, R. (2012). Building a comprehensive R&D community on serious games. *Procedia Computer Science*, *15*, 1–3. doi:10.1016/j.procs.2012.10.051.

De Vaus, D. (2001). Research design in social research. *British Educational Research Journal*, *28*, 296. Sage. Retrieved May 3, 2013, from http://www.amazon.com/Research-Design-Social-David-Vaus/dp/0761953477

DiGRA—Digital Games Research Association (n.d.). Retrieved December 20, 2012, from http://www.digra.org/

Duke, R. D. (1974). *Gaming: the future's language* (1st ed., p. 223). New York: Sage Publications.

Duke, R. D. (1981). Slogan. (Simulation game for team training), Simulation Experience. Retrieved May 3, 2013, from http://www.simxp.com/en/simulations-games/slogan/

Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*, 2, 350. doi:10.2307/2666999.

Egenfeldt-Nielsen, S. (2005). *Beyond edutainment: exploring the educational potential of computer games. department of digital aesthetics & communication.* Copenhagen: IT-University of Copenhagen. Retrieved May 3, 2013, from http://www.seriousgames.dk/downloads/egenfeldt.pdf

Egenfeldt-Nielsen, S. (2006). Overview of research on the educational use of video games. *Digital Kompetanse*, *1*, 184–213. Retrieved May 3, 2013, from http://www.idunn.no/ts/dk/2006/03/overview_of_research_on_the_educationaluseof_video_games?languageId=2

Enochsson, L., Isaksson, B., Tour, R., Kjellin, A., Hedman, L., Wredmark, T. *et al* (2004). Visuospatial skills and computer game experience influence the performance of virtual endoscopy. *Journal of Gastrointestinal Surgery Official Journal of the Society for Surgery of the Alimentary Tract*, *8*, 7, 876–882; discussion 882. Retrieved May 3, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/15531242

Erfani, M., El-Nasr, M., Milam, D., Aghabeigi, B., Lameman, B., Riecke, B. E. *et al* (2010). The effect of age, gender, and previous gaming experience on game play performance. In P. Forbrig, F. Paternó & A. Mark Pejtersen (Eds), *Human-computer interaction* Vol. 332 (pp. 293–296). Boston: Springer. doi:10.1007/978-3-642-15231-3_33.

Ermi, L. & Mäyrä, F. (2003). Power and control of games: children as the actors of game cultures. *Digital Games Research Conference (Digra)* (pp. 234–244). University of Utrecht. Retrieved May 3, 2013, from http://www.uta.fi/~tlilma/Ermi_Mayra_Power_and_Control_of_Games.pdf

Ferris, G. R. (2005). Development and validation of the political skill inventory. *Journal of Management*, *31*, 1, 126–152. doi:10.1177/0149206304271386.

GALA—Network of Excellence for Serious Games (n.d.). Retrieved May 3, 2013, from http://www.galanoe.eu/

Garris, R., Ahlers, R. & Driskell, J. E. (2002). Games, motivation, and learning: a research and practice model. *Simulation & Gaming*, *33*, 4, 441–467. doi:10.1177/1046878102238607.

Geurts, J. L. A., Duke, R. D. & Vermeulen, P. A. M. (2007). Policy gaming for strategy and change. *Long Range Planning*, *40*, 6, 535–558. doi:10.1016/j.lrp.2007.07.004.

Girard, C., Ecalle, J. & Magnan, A. (2012). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, *19*, 3, 207–219. doi:10.1111/j.1365-2729.2012.00489.x.

Good, B. M. & Su, A. I. (2011). Games with a scientific purpose. *Genome Biology*, *12*, 12, 135. doi:10.1186/gb-2011-12-12-135.

Gosen, J. & Washbush, J. (2004). A review of scholarship on assessing experiential learning effectiveness. *Simulation & Gaming*, *35*, 2, 270–293. doi:10.1177/1046878104263544.

Greenblat, C. S. (1973). Teaching with simulation games: a review of claims and evidence. (C. Greenblat & R. D. Duke, Eds.) *Teaching Sociology*, *1*, 1, 62–83.

Guay, F., Vallerand, R. J. & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: the Situational Motivation Scale (SIMS). *Motivation and Emotion*, *24*, 3, 175–213.

Hainey, T. (2010). *Using games-based learning to teach requirements collection and analysis at tertiary education level*. Edinburgh: University of the West of Scotland. Retrieved May 3, 2013, from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Using+Games-Based+Learning+to+Teach+Requirements+Collection+and+Analysis+at+Tertiary+Education+Level#1

Hainey, T. & Connolly, T. M. (2010). Evaluating games-based learning. *International Journal of Virtual and Personal Learning Environments*, *1*, 1, 57–71. doi:10.4018/jvple.2010091705.

Harper, J. D., Kaiser, S., Ebrahimi, K., Lamberton, G. R., Hadley, H. R., Ruckle, H. C. *et al* (2007). Prior video game exposure does not enhance robotic surgical performance. *Journal of Endourology / Endourological Society*, *21*, 10, 1207–1210. doi:10.1089/end.2007.9905.

Harteveld, C. (2011). *Triadic game design*. London: Springer. Retrieved May 3, 2013, from http://books.google.nl/books?id=sIcQYKcoH7IC

Harteveld, C. (2012). *Making sense of virtual risks; a quasi experimental investigation into game-based training*. (p. 435). Delft, The Netherlands: IOS Press. doi: 10.4233/uuid:c3d38ad9-88cb-403c-b69e-ba22b005a26f.

Hays, R. T. (2005). *The effectiveness of instructional games: a literature review and discussion*. Orlando: Naval War Center: Training Systems Division.

Hofstede, G. (1986). Cultural differences in teaching and learning. (G. Hofstede, Ed.) *International Journal of Intercultural Relations*, *10*, 3, 301–320. doi:10.1016/0147-1767(86)90015-5.

Hofstede, G. J., De Caluwe, L. & Peters, V. A. M. (2010). Why simulation games work-in search of the active substance: a synthesis. *Simulation & Gaming*, *41*, 6, 824–843. doi:10.1177/1046878110375596.

Holsbrink-Engels, G. A. (1998). *Computer-based role playing for interpersonal skills training*. Enschede: Universiteit Twente.

Huang, W.-H. (2011). Evaluating learners' motivational and cognitive processing in an online game-based learning environment. *Computers in Human Behaviour*, *27*, 2, 694–704. doi:10.1016/j.chb.2010.07.021.

ISAGA (n.d.). Retrieved May 3, 2013, from http://www.isaga.com/

Jackson, C. L., Colquitt, J., Wesson, M. J. & Zapata-Phelan, C. P. (2006). Psychological collectivism: a measurement validation and linkage to group member performance. *The Journal of Applied Psychology*, *91*, 4, 884–899. doi:10.1037/0021-9010.91.4.884.

Jenson, J. & De Castell, S. (2010). Gender, simulation, and gaming: research review and redirections. *Simulation & Gaming*, *41*, 1, 51–71. doi:10.1177/1046878109353473.

Kato, P. M., Cole, S. W., Bradlyn, A. S. & Pollock, B. H. (2008). A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial. *Pediatrics*, *122*, 2, e305–e317. doi:10.1542/peds.2007-3134.

Ke, F. (2009). A qualitative meta-analysis of computer games as learning tools (chapter 1) (pp. 1–33). IGI.

Kenny, R. F. & Gunter, G. (2011). Factors affecting adoption of video games in the classroom. *Journal of Interactive Learning Research*, *22*, 2, 259–276. Retrieved May 3, 2013, from http://www.editlib.org/p/32258

Kenny, R. F. & McDaniel, R. (2011). The role teachers' expectations and value assessments of video games play in their adopting and integrating them into their classrooms. *British Journal of Educational Technology*, *42*, 2, 197–213. doi:10.1111/j.1467-8535.2009.01007.x.

Kickmeier-Rust, M. D., Steiner, C. M. & Albert, D. (2009). Non-invasive assessment and adaptive interventions in learning games. *Intelligent Networking and Collaborative Systems 2009 INCOS 09 International Conference on* (pp. 301–305). IEEE. doi:10.1109/INCOS.2009.30.

Kinzie, M. B. & Joseph, D. R. D. (2008). Gender differences in game activity preferences of middle school children: implications for educational game design. *Educational Technology Research & Development*, *56*, 5–6, 643–663. doi:10.1007/s11423-007-9076-z.

Kirkpatrick, D. L. (1994). *Evaluating training programs: the four levels*. San Francisco: Berret-Koehler.

Kirkpatrick, D. L. (2006). Seven keys to unlock the four levels of evaluation. *Performance Improvement*, *45*, 7, 5–8.

Knight, J. F., Carley, S., Tregunna, B., Jarvis, S., Smithies, R., De Freitas, S. *et al* (2010). Serious gaming technology in major incident triage training: a pragmatic controlled trial. *Resuscitation*, *81*, 9, 1175–1179. Retrieved May 3, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/20732609

Kortmann, R., Bekebrede, G., Van Daalen, C. E., Harteveld, C., Mayer, I. S. & Van Dierendonck, D. (2012). Veerkracht—a game for servant-leadership development. *Proceedings of the 43rd annual conference of the International Simulation and Gaming Association*.

Kriz, W. C. & Hense, J. U. (2004). Evaluation of the EU-project "Simgame" in business education. In W. C. Kriz & T. Eberle (Eds), *Bridging the gap: transforming knowledge into action through gaming and simulation (Isaga 2004 Conference Proceedings)* (pp. 352–363). Munchen: Sagsaga.

Kriz, W. C. & Hense, J. U. (2006). Theory-oriented evaluation for the design of and research in gaming and simulation. *Simulation & Gaming*, *37*, 2, 268–283. doi:10.1177/1046878106287950.

Kuit, M. (2002). Strategic behavior and regulatory styles in the Netherlands Energy Industry. Retrieved May 3, 2013, from http://www.narcis.nl/publication/RecordID/oai:tudelft.nl:uuid:f940a515-f967-4cf3-9a96-07cb69a61ae5

Lee, J. (1999). Effectiveness of computer-based instructional simulation: a meta analysis. *International Journal of Instructional Media*, *26*, 1, 71–85. Retrieved May 3, 2013, from http://www.questia.com/googleScholar.qst?docId=5001238108

Lee, K. & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*, 2, 329–358. doi:10.1207/s15327906mbr3902_8.

Leemkuil, H. (2006). *Is it all in the game? Learner support in an educational knowledge management simulation game*. Enschede, The Netherlands: University of Twente. Retrieved May 3, 2013, from http://users.gw.utwente.nl/leemkuil/PhDThesisLeemkuil2006.pdf

Lepper, M. & Malone, T. W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. In R. E. Snow & M. J. Farr (Eds), *Aptitude, learning, and instruction. III. Cognitive and affective process analysis* (pp. 255–286). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Leyton-Brown, K. & Shoham, Y. (2008). Essentials of game theory. In R. J. Brachman & T. Dietterich (Eds), *Political science* Vol. 2 (p. 104). San Francisco, CA: Morgan & Claypool Publishers. doi:10.2200/S00108ED1V01Y200802AIM003.

Ma, Y., Williams, D., Prejean, L. & Richard, C. (2007). A research agenda for developing and implementing educational computer games: colloquium. *British Journal of Educational Technology*, *38*, 3, 513–518.

Mackenzie, N. & Knipe, S. (2006). Research dilemmas: paradigms, methods and methodology. *Issues In Educational Research*, *16*, 2, 193–205. Retrieved May 3, 2013, from http://www.iier.org.au/iier16/mackenzie.html

Malone, T. W. & Lepper, M. (1987). Making learning fun: a taxonomy of intrinsic motivation for learning. In R. E. Snow & M. J. Farr (Eds), *Aptitude, learning, and instruction: III. Cognitive and affective process analysis* (pp. 223–225). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Martin, A. J. & Jackson, S. (2008). Brief approaches to assessing task absorption and enhanced subjective experience: examining "short" and "core" flow in diverse performance domains. *Motivation and Emotion*, *32*, 3, 141–157. doi:10.1007/s11031-008-9094-0.

Mayer, I. S. (2009). The gaming of policy and the politics of gaming: a review. (Y. G. Kin, C. Yiyu, & Y. Gee, Eds.) *Simulation & Gaming*, *40*, 6, 825–862. doi:10.1177/1046878109346456.

Mayer, I. S., Carton, L., De Jong, M., Leijten, M. & Dammers, E. (2004). Gaming the future of an urban network. *Futures*, *36*, 3, 311–333. doi:10.1016/S0016-3287(03)00159-9.

Mayer, I. S., Warmelink, H. J. G. & Bekebrede, G. (2013). Learning in a game-based virtual environment: a comparative evaluation in higher education. *European Journal of Engineering Education*, *38*, 1, 85–106. doi:10.1080/03043797.2012.742872.

Mayer, I. S., Zhou, Q., Lo, J., Abspoel, L., Keijser, X., Olsen, E. *et al* (2012). Integrated, ecosystem-based marine spatial planning?: First results from international simulation-game experiment. *Third international engineering systems symposium CESUN 2012, Delft University of Technology, 18–20 June 2012* (pp. 18–20). Delft, The Netherlands: IEEE.

Mayer, I. S., Zhou, Q., Lo, J., Abspoel, L., Keijser, X., Olsen, E. *et al* (in press). Integrated, ecosystem-based marine spatial planning: design and results of a game-based quasi-experiment. *Ocean and Coastal Management*. doi:10.1016/j.ocecoaman.2013.04.006

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge, MA: Cambridge University Press. Retrieved May 3, 2013, from http://books.google.com/books?hl=en&lr=&id=duWx8fxkkk0C&oi=fnd&pg=PA31&dq=Cognitive+Theory+of+Multimedia+Learning&ots=x62cq2Ag4C&sig=km1JhOxQR2OGZunazZVGY6qDU3Q

Mayes, D. K. & Cotton, J. E. (2001). Measuring engagement in video games: a questionnaire. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 692–696). Retrieved May 3, 2013,

from http://www.scopus.com/inward/record.url?eid=2-s2.0-0442310948&partnerID=40&md5=b56e3e59c17f58d4d21a35d7755b9f6a

McGonigal, J. E. (2011). *Reality is broken: why games make us better and how they can change the world.* Vol. 22 Penguin Books, Ed. New York: The Penguin Press HC. (p. 400). Retrieved May 3, 2013, from http://www.amazon.com/dp/1594202850

Meyer, B. (2010). Comparative studies in game-based language learning: a discussion of methodology. (P. Escudeiro, Ed.) *Proceedings of the 9th European Conference on ELearning Vol 1* (pp. 362–369).

Mortagy, Y. & Boghikian-whitby, S. (2010). A longitudinal comparative study of student perceptions in online education. *Interdisciplinary Journal of Elearning and Learning Objects*, 6, 23–44. Retrieved May 3, 2013, from http://ijello.org/Volume6/IJELLOv6p023-044Mortagy684.pdf

Noy, A., Raban, D. R. & Ravid, G. (2006). Testing social theories in computer-mediated communication through gaming and simulation. *Simulation & Gaming*, 37, 2, 174–194. doi:10.1177/1046878105286184.

Olsen, T., Procci, K. & Bowers, C. (2011). Serious games usability testing: how to ensure proper usability, playability, and effectiveness. (A. Marcus, Ed.) *Design User Experience and Usability Theory Methods Tools and Practice Proceedings First International Conference DUXU 2011 Held as Part of HCI International 2011*, 6770, 625–634. doi:10.1007/978-3-642-21708-1.

Oslin, J. L., Mitchell, S. A. & Griffin, L. L. (1998). The Game Performance Assessment Instrument (GPAI): some concerns and solutions for further development. *Journal of Teaching in Physical Education*, 17, 2, 220–240. Retrieved May 3, 2013, from http://www.cababstractsplus.org/abstracts/Abstract.aspx?AcNo=19981804818

Paas, F., Tuovinen, J. E., Merriënboer, J. J. G. & Aubteen Darabi, A. (2005). A motivational perspective on the relation between mental effort and performance: optimizing learner involvement in instruction. *Educational Technology Research and Development*, 53, 3, 25–34. doi:10.1007/BF02504795.

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: impact on educational effectiveness and student motivation. *Computers & Education*, 52, 1, 1–12. doi:10.1016/j.compedu.2008.06.004.

Pfister, R. (2011). Gender effects in gaming research: a case for regression residuals? *Cyberpsychology Behavior and Social Networking*, 14, 10, 603–607. doi:10.1089/cyber.2010.0547.

Preschl, B., Wagner, B., Forstmeier, S. & Maercker, A. (2011). E-health interventions for depression, anxiety disorder, dementia, and other disorders in older adults: a review. *Cyber Therapy and Rehabilitation*, 4, 3, 371–385. Retrieved May 3, 2013, from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc7&NEWS=N&AN=2011-25383-006

Randel, J. M., Morris, B. A., Wetzelf, C. D. & Whitehill, B. V. (1992). The effectiveness of games for educational purposes: a review of the research. *Simulation & Gaming*, 25, 261–276. Retrieved May 3, 2013, from http://www.cofc.edu/~seay/cb/findsim.html

Raphael, C., Bachen, C., Lynn, K. M., Baldwin-Philippi, J. & McKee, K. A. (2009). Games for civic learning: a conceptual framework and agenda for research and design. *Games and Culture*, 5, 2, 199–235. doi:10.1177/1555412009354728.

Reichlin, L., Mani, N., McArthur, K., Harris, A. M., Rajan, N. & Dacso, C. C. (2011). Assessing the acceptability and usability of an interactive serious game in aiding treatment decisions for patients with localized prostate cancer. *Journal of Medical Internet Research*, 13, 1, e4. doi:10.2196/jmir.1519.

Salomon, M. & Soudoplatoff, S. (2010). Why virtual-world economies matter. *Journal of Virtual Worlds Research*, 2, 4, 14. Retrieved May 3, 2013, from http://journals.tdl.org/jvwr/article/viewArticle/859

Sanchez, A., Cannon-bowers, J. & Bowers, C. (2010). Establishing a science of game based learning. In J. A. Cannon-Bowers & C. Bowers (Eds.), *Serious game design and development: technologies for training and learning* (pp. 290–304). Hershey, PA: IGI Global. doi:10.4018/978-1-61520-739-8.ch016.

Scharpf, F. W. (1997). Games real actors play?: actor-centered institutionalism in policy research. In P. Sabatier (Ed.), *Theoretical lenses on public policy* (p. 318). Boulder, CO: Westview Press. Retrieved May 3, 2013, from http://opac.rero.ch/get_bib_record.cgi?db=ne&rero_id=R236641760

Schneider, D. K. (2005). Research design for educational technologists. Geneva: AACE. [unpublished working document]. Retrieved May 3, 2013, from http://www.editlib.org/index.cfm?fuseaction=Reader.ViewAbstract&paper_id=28664

Schuurink, E., Houtkamp, J. & Toet, A. (2008). Engagement and EMG in serious gaming: experimenting with sound and dynamics in the Levee Patroller training game. In P. Markopoulos, B. de Ruyter, W. IJsselsteijn & D. Rowland (Eds.), *Fun and games. Second International Conference, Eindhoven, The Netherlands, October 20–21, 2008. Proceedings* (pp. 139–149). Berlin, Heidelberg: Springer Verlag. doi:10.1007/978-3-540-88322-7_14.

SG Academy (n.d.). Retrieved December 20, 2012, from http://academy.seriousgamessociety.org/

SGA—Serious Games Association (n.d.). Retrieved May 3, 2013, from http://www.seriousgamesassociation.com/

Shaffer, D. W. (2006). Epistemic frames for epistemic games. *Computers & Education*, 46, 3, 223–234. doi:10.1016/j.compedu.2005.11.003.

Sharkworld (n.d.). *(Serious game about project management)*. Rotterdam: Ranj. Retrieved May 3, 2013, from http://www.sharkworld.nl/

Shubik, M. (1999). *Political economy, oligopoly and experimental games: the selected essays of Martin Shubik volume one*. Cheltenham: Elgar.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds), *Computer games and instruction* Vol. 55 (pp. 503–524). Charlotte, NC: Information Age Publishers. Retrieved May 3, 2013, from http://proxygw.wrlc.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2011-11269-020&site=eds-live&scope=site

Shute, V. J., Masduki, I. & Donmez, O. (2010). Conceptual framework for modeling, assessing and supporting competencies within game environments. *Technology Instruction Cognition and Learning*, 8, 2, 137–161. Retrieved May 3, 2013, from http://myweb.fsu.edu/vshute/pdf/TICL2010.pdf

Shute, V. J., Ventura, M., Bauer, M. & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning?: Flow and grow. (U. Ritterfeld, M. Cody, & P. Vorderer, Eds.) *Serious Games Mechanisms and Effects*, 1, 1, 1–33. Retrieved May 3, 2013, from http://books.google.com/books?hl=en&lr=&id=GwPf7tbO5mgC&pgis=1

SprintCity (n.d.). *(Serious game about urban planning)*. Rotterdam, The Netherlands: Vereniging Deltametropool. Retrieved May 3, 2013, from http://deltametropool.nl/nl/sprintstad_in_utrecht

Squire, K. (2002). Cultural framing of computer/video games. *Game Studies. The International Journal of Computer Game Research*, 2, 1. Retrieved May 3, 2013, from http://www.gamestudies.org/0102/squire/

Squire, K. (2004). *Replaying history: learning world history through playing civilization III*. Bloomington, IN: Indiana University. Retrieved May 3, 2013, from http://website.education.wisc.edu/kdsquire/dissertation.htm

Steunenberg, B., Schmidtchen, D. & Koboldt, C. (1999). Strategic power in the european union: evaluating the distribution of power in policy games. *Journal of Theoretical Politics*, 11, 3, 339–366. doi:10.1177/0951692899011003005.

Stichting Rechten Online (2007). Cyberdam. Stichting Rechten Online. Retrieved August 10, 2012, from http://www.cyberdam.nl

Szturm, T., Betker, A. L., Moussavi, Z., Desai, A. & Goodman, V. (2011). Effects of an interactive computer game exercise regimen on balance impairment in frail community-dwelling older adults: a randomized controlled trial. *Physical Therapy*, 91, 10, 1449–1462. doi:10.2522/ptj.20090205.

Tallir, I. B., Lenoir, M., Valcke, M. & Musch, E. (2007). Do alternative instructional approaches result in different game performance learning outcomes? Authentic assessment in varying game conditions. *International Journal of Sport Psychology*, 38, 3, 263–282.

Trepte, S. & Reinecke, L. (2011). The pleasures of success: game-related efficacy experiences as a mediator between player performance and game enjoyment. *Cyberpsychology Behavior and Social Networking*, 14, 9, 555–557. Retrieved May 3, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/21342012

TU Delft, Tygron Serious Gaming, & Port of Rotterdam (n.d.). SimPort MV2 serious game. Delft, The Netherlands: Tygron serious games and media. Retrieved May 3, 2013, from http://www.simport.eu/

Tykhonov, D., Jonker, C. & Meijer, S. A. (2008). Agent-based simulation of the trust and tracing game for supply chains and networks. *and Social Simulation*. Retrieved May 3, 2013, from http://jasss.soc.surrey.ac.uk/11/3/1.html

Van der Spek, E. D. (2011). (October 26). *Experiments in serious game design: a cognitive approach. SIKS Dissertation Series*. P. Retrieved May 3, 2013, from http://www.narcis.nl/publication/RecordID/oai:dspace.library.uu.nl:1874/211480

Van der Spek, E. D., Wouters, P. & Van Oostendorp, H. (2011). Code red: triage or cognition-based design rules enhancing decisionmaking training in a game environment. *British Journal of Educational Technology*, 42, 3, 441–455. doi:10.1111/j.1467-8535.2009.01021.x.

Vartiainen, P. (2000). Evaluation methods and comparative study. *European Evaluation Society EES Conference* (pp. 12–14). Retrieved May 3, 2013, from http://cjpe.ca/distribution/20001012_vartiainen_pirkko.pdf

Ventum Online (n.d.). *(Serious project management game)*. Delft: TU Delft, Faculteit TBM. Retrieved May 3, 2013, from http://www.ventumonline.nl/

Venture Simulations Ltd. (n.d.). *Simventure. (serious business game)*. York, UK: Venture Simulations Ltd. Retrieved May 3, 2013, from http://www.simventure.co.uk/

Verdaasdonk, E. G. G., Dankelman, J., Schijven, M. P., Lange, J. F., Wentink, M. & Stassen, L. P. S. (2009). Serious gaming and voluntary laparoscopic skills training: a multicenter study. *Minimally Invasive Therapy & Allied Technologies?: MITAT?: Official Journal of the Society for Minimally Invasive Therapy*, *18*, 4, 232–238. doi:10.1080/13645700903054046.

Wolfe, J. & Box, T. M. (1988). Team cohesion effects on business game performance. *Simulation & Gaming*, *19*, 1, 82–98. doi:10.1177/003755008801900105.

Wood, R. T. A., Griffiths, M. D. & Eatough, V. (2004). Online data collection from video game players: methodological issues. *CyberPsychology Behavior*, *7*, 5, 511–518. doi:10.1089/cpb.2004.7.511.

Wright, J. D. & Marsden, P. V. (Eds.) (2010). *Handbook of survey research* (2nd ed.). Bingley, UK: Emerald Group Publishing.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B. *et al* (2012). Our princess is in another castle: a review of trends in serious gaming for education. *Review of Educational Research*, *82*, 1, 61–89. doi:10.3102/0034654312436980.

Yusoff, A., Crowder, R. & Gilbert, L. (2010). Validation of serious games attributes using the technology acceptance model. (IEEE, Ed.) *2010 Second International Conference on Games and Virtual Worlds for Serious Applications*, *0*, 45–51. doi:10.1109/VS-GAMES.2010.745.

Zhou, Q., Mayer, I. S., Bekebrede, G., Warmerdam, J. & Knepflé, M. (2013). The climate game: connecting water management and spatial planning through simulation gaming? In J. Edelenbos, N. Bressers & P. Scholten (Eds), *Water governance as connective capacity* (pp. 109–127). Farnham, Surrey, UK: Ashgate Publishing Ltd.

Appendix

Table A1: Overview of game cases

Game name	Start year	End year	Current status	Client organization	Game concept	Game platform / engine	Game objective	Education (TU-Delft) students	Professional learning & training	Support policymaking & management	Academic research
Veerkracht (Resilience)	2011	2013	Pending	Rijkswaterstaat, the Netherlands	Custom made	Custom made	Leadership and personal competences training		✓	✓	✓
MSP (Marine Spatial Planning) Challenge 2011	2011	Continues	Pending	Ministry Infrastructure & Environment, the Netherlands. Helcom, Ospar, ICES	Custom made	Custom made	Insight in complexity Marine Spatial Planning		✓	✓	✓
Shark World	2010	Continues	Used occasionally	TU-Delft	COTS	COTS	Training project management skills	✓			
Sprintstad	2009	Continues	Used frequently by client	Vereniging Deltametropool,	Custom made	Custom made	Insight in complexity of urban planning and transportation	✓		✓	
Team-Up	2009	Continues	Used occasionally		Custom made	Mod	Team role performance and competences training and assessment.	✓	✓		✓
TopSim	2009	Continues	Used occasionally	TU-Delft	COTS	COTS	Business and personal competences training	✓			
Supervisor / Hazard Recognition	2008	2011	Research	Royal Dutch Shell, corporate learning	Custom made	Mod	Training of operational safety and emergency procedures		✓		✓
Construct.IT	2008	2009	Closed	3TU construction management	Custom made	Custom made	Insight and skills for construction management projects	✓			
Watergame	2006	Continues	Used frequently by client		MOD	Custom made	Insight in complexity of water management, urban planning	✓	✓	✓	✓
Cyberdam 3D	2006	2007	Closed	n.a.	Custom made	COTS	Training skills and theory on policy making	✓			
Sim Port-MV2	2005	Continues	Used frequently by client	Port of Rotterdam (PoR)	Custom made	Custom made	Insight in complexity of port planning	✓	✓	✓	✓
Levee Patroller	2005	Continues	Used actively by client	Deltares, Water Boards	Custom made	Mod	Training of operational safety and emergency procedures	✓	✓		✓
Sim-MV2	2004	2005	Closed	Port of Rotterdam	Custom made	Custom made	Insight in complexity of port planning	✓	✓	✓	✓
Sieberdam : railway area reconstruction	2004	2009	Closed	n.a.	Custom made	COTS	Training skills and theory on policy making	✓			✓
Ventum Online	2003	2012	Used occasionally	n.a.	Custom made	Custom made	Training skills and theory on project management	✓	✓		

Table A1: Continued

Game name	Virtual	Technology use	Stand alone / networked	Single / multiplayer	External realism	Player interaction	Paraphernalia	Min no players per game instance	Optimal no of participants in session	Average (un)interrupted duration of game in hours
Veerkracht (Resilience)	2D visuals	Multimedia supported	Networked	Multiplayer	High	Role play	Game manual; game maps; game objects; game software; computers;	15	15–30	8
MSP (Marine Spatial Planning) Challenge 2011	2D virtual	Computer supported	Stand alone	Multiplayer	High	Role play	Game manual; game maps; game objects; game software; computers;	100		8
Shark World	2D virtual	Multimedia supported	Networked	Single player	Low	Competition	Game manual; game software; computers; Internet	1		4
Sprintstad	2D virtual	Computer supported	Networked	Multiplayer	High	Role play	Game manual; game maps; game objects; game software; computers; Internet	6		6
Team-Up	3D virtual	Computer supported	Networked	Multiplayer	Low	Cooperation; competition	Game manual; game maps; game objects; game software; computers; Internet	4		1
TopSim	2D visuals	Computer supported	Networked	Multiplayer	Medium	Competition	Game manual; game maps; game objects; game software; computers; Internet	?		?
Supervisor / Hazard Recognition	3D virtual	Computer-based	Stand alone	Single player	High	Action	Game manual; game maps; game objects; game software; computers; Internet	1		2
Construct.IT	2.5D virtual	Computer supported	Networked	Multiplayer	High	Role play	Game manual; game maps; game objects; game software; computers; Internet	15		8
Watergame	2.5D virtual	Computer based	Networked	Multiplayer	High	Role play	Computers and Internet; game software	4		8
Cyberdam 3D	3D virtual	Computer based	Networked	Multiplayer	Medium	Role play	Game manual; game maps; game objects; game software; computers; Internet	?		n.a. online
Sim Port-MV2	2.5D visuals	Computer-based	Networked	Multiplayer	High	Role play	Computers and Internet; game software	5	10–15 (2/3 teams)	7
Levee Patroller	3D virtual	Computer based	Stand alone	Single player	High	Action	Computers and Internet; @@game software	1		4
Sim-MV2	2D virtual	Computer based	Networked	Multiplayer	High	Role play	Game manual; game maps; game objects; game software; computers; Internet	5	10–15 (2/3 teams)	7
Sieberdam : railway area reconstruction	2D visuals	Computer supported	Networked	Multiplayer	Medium	Role play	Game manual; game maps; game objects; game soft ware; computers; Internet	20		n.a. online
Ventum Online	2D visuals	Computer-based	Networked	Multiplayer	Medium	Role play	Game manual; game maps; game objects; game software; computers; Internet	8	12	7

Table A2: Dataset (on January 10, 2013)

Game name/characteristics	1. Simport	2. Ventum online	3. Cyberdam	4. Construct.it	5. Climate game	6. SprintCity	7. SharkWorld	8. TeamUp	9. MSP	10. Slogan	11. SimVenture	12. Servant leadership	Total freq.	Freq in % subcategory
n = freq. in row	482	683	635	65	36	45	23	347	96	26	26	24	2488	100
n = % of total row	19.4%	27.3%	25.4%	2.6%	1.4%	1.8%	0.9%	13.9%	3.8%	1.0%	1.0%	1.0%		100
Earliest date of play	31-8-2005	14-10-2005	21-2-2006	7-1-2009	11-6-2010	17-12-2009	15-12-2010	25-11-2010	3-11-2011	20-6-2012	20-6-2012	20-4-2012		
Latest date of play	13-3-2012	5-6-2012	20-4-2012	6-1-2010	24-11-2011	22-6-2010	15-12-2010	29-11-2012	31-10-2012	20-6-2012	20-6-2012	20-4-2012		
Number of game sessions (days with one or more instances)														
Freq.	34	23	49	3	3	8	1	19	2	1	1	1	145	100
%	23.5	15.9	33.8	2.1	2.1	5.5	0.7	13.1	1.4	0.7	0.7	0.7		100
No of indep. game instances (groups)	100	70	49	3	9	10	23	61	2	2	13	1	322	100
Nationality														
Netherlands	—	—	—	—	—	—	—	29	18	14	14	24	99	4.0
Other	12	2	29	—	—	—	—	27	67	12	12	n.a.	161	6.5
Missing	470	681	606	65	36	45	23	291	11	0	0	0	2228	89.6
Age														
Age.y mean (SD)	27.6 (8.9)	23.7 (1.9)	21.3 (4.0)	23.9 (2.5)	40.2 (12.8)	37 (9.2)	21.2 (1.6)	35.7 (9.8)	41.7 (11.0)	24.8 (2.1)	24.8 (2.1)	45.8 (7.7)		100
16–20	35	1	266	1	0	0	8	0	0	0	0	0	311	16.5
21–25	203	449	173	40	7	0	12	43	4	17	17	0	965	51.3
26–30	65	60	22	11	2	3	0	44	9	9	9	0	234	12.4
31–35	28	4	4	0	0	1	0	36	21	0	0	1	95	5.1
36–40	21	1	3	0	4	1	0	33	10	0	0	5	78	4.1
41–45	13	0	4	0	5	1	0	26	10	0	0	7	66	3.5
46–50	10	0	1	0	7	2	0	23	13	0	0	5	61	3.2
>50	17	0	1	0	4	0	0	24	21	0	0	6	73	3.9
Gender													2280	*100*
Male	355	510	299	53	24	5	20	279	44	22	22	17	1650	72.4
Female	103	161	237	12	8	3	3	50	40	4	4	5	630	27.6
Student/professional													2428	*100*
Student	326	662	632	65	10	0	23	61	0	26	26	0	1831	75.4
Professional	136	19	0	0	26	45	0	284	63	0	0	24	597	24.6
Level of education													2190	*100*
Secondary education	0	0	0	0	—	—	0	—	—	0	0	—	0	0.0
Middle vocational education	9	0	210	0	—	—	0	—	—	0	0	—	219	6.0
Higher vocational education	155	0	345	0	—	—	0	—	—	0	0	—	500	13.6
University	126	662	555	63	—	—	23	27	—	26	26	15	1471	39.9

Table A2: Continued

Game name/characteristics	1. Simport	2. Ventum online	3. Cyberdam	4. Construct.it	5. Climate game	6. SprintCity	7. SharkWorld	8. TeamUp	9. MSP	10. Slogan	11. SimVenture	12. Servant leadership	Total freq.	Freq in % subcategory
Playing analogue games													*1493*	*100*
1. Rarely	51	47	62	11	1	—	1	—	17	—	—	—	190	12.7
2. Couple of times a year	131	225	224	29	7	—	12	—	40	—	—	—	668	44.7
3. Monthly	79	155	128	11	4	—	5	—	9	—	—	—	391	26.2
4. Weekly	37	97	54	6	4	—	4	—	7	—	—	—	209	14.0
5. daily	7	18	8	1	0	—	0	—	1	—	—	—	35	2.3
Playing digital games													*1853*	*100.0*
1. Rarely	53	90	81	11	14	3	1	82	33	2	2	—	372	20.1
2. Couple of times a year	98	154	139	21	8	2	6	81	27	13	13	—	562	30.3
3. Monthly	55	110	95	8	4	2	4	42	6	9	9	—	344	18.6
4. Weekly	63	139	114	12	6	1	7	46	6	3	3	—	400	21.6
5. Daily	36	49	52	6	0	0	4	20	6	1	1	—	175	9.4
Game profile (example constructs)														
Construct 1: Clarity of the SG (5 items)	3.8 (0.6)	3.4 (0.6)	3.3 (0.7)	3.7 (0.5)	3.9 (0.9)	3.5 (0.5)	3.2 (0.7)	3.8 (0.9)	4.1 (0.6)	3.4 (0.8)	3.6(0.8)	4.2 (0.5)	Cronbach's α = 0.76	n = 1170
Construct 2: Realism of the SG (2 items)	3.7 (0.7)	3.5 (0.7)	—	3.2 (0.8)	4.1 (0.7)	3.3 (0.9)	2.7 (0.8)	4.1 (0.8)	4.0 (0.8)	3.5 (0.9)	4.1 (1.1)	3.6 (1.1)	Cronbach's α = 0.71	n = 1277
Construct 3: Learning satisfaction (var. items)	3.0 (0.8)	3.2 (0.6)	3.2 (0.6)	3.1 (0.7)	4.1 (0.9)	3.4 (1.1)	2.8 (07)	n.a.	3.6 (0.8)	4.0 (0.5)	4.0 (0.5)	3.2 (0.5)	Cronbach's α = 0.88–0.91	n = 1151
Construct 4: Team Engagement (3 items)	3.8 (0.7)	3.6 (0.7)	3.3 (0.9)	3.6 (0.7)	3.5 (0.8)	3.4 (0.9)	—	4.2 (0.6)	—	3.5 (1.0)	4.2 (0.7)	—	Cronbach's α = 0.74	n = 1145
Construct 5 Facilitator quality (2 items)	4.0 (0.6)	3.8 (0.7)	3.4 (0.9)	3.5 (0.7)	4.1 (0.8)	4.0 (0.5)	2.6 (0.8)	4.1 (0.9)	4.1 (0.9)	4.3 (0.6)	3.6 (0.9)	3.5 (0.9)	Cronbach's α = 0.76	n = 1170
Construct 6: User interaction (2 items)	3.7 (0.6)	3.7 (0.7)	—	3.4 (0.7)	3.8 (0.9)	3.3 (0.9)	3.5 (0.8)	4.1 (0.7)	3.8 (0.7)	n.a.	3.2 (0.8)	—	Cronbach's α = 0.69	n = 1634
Construct 7: Identification with role	3.9 (0.6)	3.7 (0.6)	3.6 (0.8)	3.9 (0.6)	3.5 (0.8)	—	3.3 (0.8)	4.1 (0.7)	—	3.6 (0.9)	4 (0.8)	4.2 (0.5)	Cronbach's α = 0.7	n = 1181