EVALUATION OF SERIOUS BOARD GAMES

Master Thesis Florine Kroes

Institute for
Sustainable
Process Technology

WAGENINGEN
UNIVERSITY & RESEARCH

“*Homo ludens*, de spelende mensch”

Johan Huizinga, 1938

MSc Thesis: Florine Kroes
Registration number: 931207479020
Date: 30-06-2019
Study programme: MSc Food Technology
Specialization: Food Innovation & Management
Chair group: Business Management & Organization
Supervisors: dr. R (Renate) Wesselink & prof. dr. WA (Wilfred) Dolfsma
Examiner a: dr. R (Renate) Wesselink
Examiner b: prof. dr. W.A. (Wilfred) Dolfsma
Course code: MST-80436

Abstract

The food industry is continuously facing complex sustainability issues calling for new ways of developing, producing, and marketing food products. To overcome these challenges businesses are forced to cooperate with the government, society, and other companies. Serious board games gained credibility to be a powerful tool for business environments to support social interaction between stakeholders within an innovation project and increase learning about sustainability problems to overcome complex challenges. However, a standardized evaluation method for serious board games in a business environment is lacking. This study aimed to investigate what contributes to the perceived quality of serious board games in a business environment to design a systematical evaluation model.
The evaluation model is based on Kirkpatrick's four level evaluation model for evaluation training in an organization: reaction, learning, transfer, and results. This model is complemented by the MEEGA+, used in the educational sector, and other models to create a complete evaluation model. The final model consists of the three levels of Kirkpatrick: reaction, learning, and transfer; subdivided into fun, challenge, social interaction, learning goals, and transfer intention. To measure the variables in the model, a questionnaire was developed. The questionnaire contained different questions regarding the quality of serious board games. The framework can be used by game designers to verify and compare the quality of serious board games. The serious games used in this paper offer a promising tool in the cooperation of individuals and companies in the framework of complex sustainability challenges.

Keywords: serious board games, evaluation framework, complex sustainability problems, business environment, cooperation

Table of content

1. Introduction

Nowadays, companies and their employees are facing a rapidly changing business environment. Consequently, they are increasingly confronted with problems and situations with a high level of complexity. The same applies to companies in the food industry (Riedel, 2011; Bremmers, 2004). The demands of the food industry increase continuously. The world population is expected to grow from the current 7.7 billion to nearly 10 billion by 2050. The Food and Agriculture Organization of the United Nations (FAO) has suggested that global food production will need to increase by 70% relative to production in 2050 (Bruinsma, 2009; Baldwin, 2015; Worldometers.info, 2019). Due to significant amounts of energy used to produce, process and transport food, the food industry is a significant factor in climate change, water use, and pollution (Baldwin, 2015). The sustainability issues related to food range widely from greenhouse gas emissions and energy use to land use, water availability, soil quality, water quality and quantity, biodiversity losses, and chemical exposure. Threats that stem from other issues, including food quality and food security, the development of genetically modified organisms, growth hormone residues in food, etc., are of concern (Notarnicola *et al.*, 2012). Besides, the food industry faces significant risks from public criticism (Maloni *et al.*, 2006).

The growing concern of consumers regarding environmental issues, in combination with the increasing demand for quality products and processes within a competitive global environment, calls for new ways of developing, producing and marketing food products (Schmidheiny, 1992; Rana *et al.,* 2009). Therefore, the stakeholder in the traditional food chains needs a significant change in the way of behavior and thinking (Bremmers, 2004). New forms of cooperation between government, business, and society are required to fulfill the quest of consumers for safe, environmentally friendly and high-standard food products (Schmidheiny, 1992; Bremmers, 2004).

Sustainability challenges require novel tools to facilitate learning and encourage dialogue between different stakeholders so that decisions can be made from a position of shared knowledge and understanding (Medema *et al.*, 2016). In literature, games have increasingly gained credibility to be a powerful learning tool for sustainability issues and stimulates social interaction (White *et al.,* 2010; van Pelt *et al.,* 2015; Dieleman *et al.*, 2006, Klabbers, 1989, p3; Abt, 1987; Michael *et al.*, 2005; Susi *et al.*, 2007). Due to the seriousness of thoughts and problems these games are covering, the games are known as “serious games” and are defined as games with a primary purpose other than pure entertainment (Abt, 1987; Susi *et al.,* 2007).

Serious Games have a hedonic component of bonding the comprehensibility of active play with the seriousness of problems they are covering. The word 'serious' reflects to the purpose of the game, in general the reason for creating the game. And the 'play' is related to the method used to achieve the purpose (Michael *et al.*, 2005).

Proper evaluated serious games can help to increase the quality of the games and improve the image of serious games (Emmerich *et al.*, 2016). Although a lot of studies have been done on serious games, detailed information about the evaluation is lacking (Connolly *et al.,* 2009). Connolly *et al.* (2009) report that an analysis of 1400 papers of serious games reveals that only 5% apply some evaluation measurement. Petri (2016) states that in the last 20 years, only seven evaluation approaches have been encountered used in 11 studies. Based on the review of numerous studies the provided evaluations are conducted for the evaluation of digital games (e.g., computer game, mobile game, video game, etc.), generally with the use of game analytics via a non-disruptive game tracking (Calderón *et al.,* 2015). However, still in multiple studies, there is not mentioned the execution of the evaluation or the questionnaires are not provided (Calderón *et al.,* 2015).

To get insights in the effectivity of the games compared to their investment cost, the need to increase the quality of the games and the possibility to compare different games on their quality aspects, there is interest for a more large-scale application of evaluations for serious board games. One organization that has an interest in such an evaluation model is ISPT.

2. Serious Board Games used by ISPT

The Institute for Sustainable Technologies (ISPT) is an organization established in Amersfoort, The Netherlands. The ambition of ISPT is to connect stakeholders from different sectors and disciplines. The institute provides an inspirational and trusted environment to support industries, SME's, scientist and government to work together and to stimulate breakthrough innovation related to sustainability (https://www.ispt.eu/). ISPT developed different serious board games that can be played with different stakeholders in an innovation project related to sustainability. A serious board game has different aims. First, the game needs to support interaction, in such a way that the players share their knowledge and visions. Secondly, the game is to educate and create awareness about a specific topic and stimulate stakeholders to participate in sustainable innovation projects. With as overall goal, that the participants start to work together in their professional lives to overcome sustainability challenges.

An example of a serious board game ISPT developed, is the ZeroBrine game. The ZeroBrine game is a serious board game in the context of the ZeroBrine project, a project which aimed to market new technologies to support the separation of brines from water in the waste stream of a company. With the use of these technologies, the brines and water can be re-used in their own company or used by other industries. Working together with the different stakeholders in the chain would stimulate the circularity of brines and water. In the game, the different stakeholders of the project are incorporated as roles. The government, industry and technology providers need to cooperate and negotiate to overcome the scarcity of water and brines. The game gives insight into the challenge and let experience the consequences of the decisions the player takes during the game. The aim of the game is to create awareness an learn about the complex issue, support social interaction and stimulate cooperation between different stakeholders.

Since ISPT is putting a lot of effort into the development of these games they are interested in the effectivity of these serious board games. Moreover, they are interested in the important quality aspects of the serious board games and how the games can be improved in such a way that the games support social interaction and learning about a complex issue. In order to get insights about the effectivity of the game an appropriate evaluation model is needed (Emmerich *et al*., 2016). A proper evaluation model helps to market the games to persuade stakeholder to participate.

In this research, there is a focus on non-digital serious board games. ISPT has an interest in games that can be played in groups, face-to-face. Moreover, they have the advantage that there is no need to learn technical skills as that is possible needed for digital games. Many complex social systems can be useful modeled with three or more people. The advantage of playing in teams is that the game allows players to learn both from their own experiences and those of their co-players, seeing the game system from different perspectives (Ickes *et al.*, 1994; Castronova *et al.*, 2015). The target group of the games people who are active in their professional life's in the field of sustainability and innovation. The aim of the games is creating awareness and understanding of a complex innovation issue related to sustainability. In addition, the game facilitates a discussion between the participants. The discussion will help to get deeper insights into each other's working field and interests.

The research focusses on the game moment itself and the immediate reaction of the participant afterward. The elements before playing the game and the results of the game play are out of scope of the research. Considering this, the design of the game and how they apply in their professional lives what they have learned are not part of this research. Moreover, personal and motivational factors are not considered.

To conclude, much research has been done on serious games. However, a large-scale evaluation model for serious board games in a business environment is lacking. To get a deeper understanding of elements influencing the quality of the game, a literature review is done. The next chapter gives insights into the unique elements of serious games, the importance of evaluation and different evaluation models coming from the business environment and educational field.

3. Theoretical background

Serious board games are used to give an authentic learning experience and creating opportunities to share experiences (Hainey *et al*., 2011). This is due to the fact they cover elements that have a positive influence on learning processes and social interaction.

3.1. Elements of serious games

First, the games allow learning through *experiencing* problem-based situations. In this way, the players learn by doing, using the information at the same time, rather than by only listening or reading (Leemkuil *et al.*, 2012; Lieberman, 2006). Second, various (scientific) knowledge can be simplified and incorporated into the game, in such a way that the before abstract *knowledge becomes accessible* to the target group (Kriz, 2003). Third, players *enjoy* playing the game, and this increases the learning experience due to the positive emotional experience (Boyle *et al*., 2011; Lieberman, 2006; Hofstede *et al.*, 2010; Abt, 1987). As well, the games provide the opportunity for *feedback* on the decision-making which increasing learning because double-loop learning occurs (Bartunek, 2014). In addition, serious games contain an *active role-play* element wherein the participant is supposed to restructure activities, information, and behaviors of a given stakeholder in a rule-based setting (Sutcliffe, 2002). With this, serious games support *social interaction* with human-to-human interactions creating opportunities to communicate business concerns and facts of the subject, efficiently (Mayer *et al*., 2014; Abt, 1987). Persons exchanging ideas, information and experiences in order to construct personal knowledge serves as a basis for common understanding and a collective

solution to a problem (Ruohomaki, 2003; Ellis *et al.*, 2003) This supports the sharing of information, knowledge, and experiences which can be helpful to develop future needs and increases understanding of different interest of stakeholders (Ruohomaki, 2003). Lastly, the games providing a *real-life experience* since the constraints in the game are addressed from the real-world (Haug *et al.*, 2011; Susi *et al.,* 2007). The players are exposed to different conditions and scenarios of the future, showing the consequences of individual and policy decisions (Juhola *et al.*, 2013). Therefore, the game stimulates thinking about the long term and provide learning about a particular subject by the influence of different forms of uncertainty (Haug *et al.*, 2011).

As serious games are covering these elements, there is an increasing interest in these types of games in professional sustainability education (Boyle *et al.*, 2009). Successful examples are the climate change game, stressed that the game could be a useful instrument in the communication of climate change (van Pelt *et al.,* 2015). Or the game about Circular Economy, where the players were able to reconnect the game to reality, think in systems and utilize critical thinking about the subject (Whalen *et al.*, 2018).

3.2 Importance of evaluation

Evaluation is a systematic process of collecting data with an overall goal to prove the game's effectiveness and suitability with regard to its designed purpose and application context (Brown & Gerhardt, 2002; Emmerich *et al.*, 2016). In recent years, increasingly serious games and applications have been developed and consequently, the research to their effectiveness is raising too (Emmerich *et al.,* 2016). Studies demonstrate the benefits of serious games in general (Chin *et al.,* 2009; Connolly *et al.,* 2012; O'Neil *et al.,* 2005) or a specific purpose in particular, such as a tool to support learning or to induce health-related behavioral changes (DeSmet *et al.,* 2014; Girard *et al.,* 2013). As each serious game is developed with a defined purpose that goes beyond entertainment, the developers want to ensure by the evaluation that this goal is achieved (Emmerich *et al.*, 2016). In general, the review of the effectiveness of serious games is based on their purpose.

For this reason, serious games need to be especially evaluated regarding their "serious" objective. Therefore, in this thesis, the effectiveness of serious games is evaluated concerning to their purpose, out of the regular usability tests or the assessments of general player experience issues (e.g., motivation). Those aspects are interesting to study as well. However, these elements, are specific for every serious game. For the game design and evaluation of

these elements general techniques and methods can be used easily and read out in other books (Bernhaupt 2010; El-Nasr *et al.,* 2013; Schultz *et al.,* 2011; Emmerich *et al.*, 2016).

Game developers are often developing serious games about a new subject, so there exist no or little examples. For instance, in the case of developing the hydrogen game, there is not a lot of information available about the use of hydrogen in the field of transportation and no other games are yet developed in this context. Therefore, designers of serious games make use of theories, literature, intuition and personal experience while developing a serious game. Despite this is a feasible way, the intended expectations of the game need to be checked and validated. The data gained can be used to improve the future game design and convince diverse stakeholders of the quality of the game. The game designers can learn more about the relationships between game design elements and the resulting player experience and consequently gain insights into the impact of games in general. As the developers need to know whether or not the player has learned the content of the serious game, the evaluation should cover the quality elements influence learning (Loh *et al.,* 2015; Emmerich *et al.,* 2016).

Thus, the evaluation of serious games is from importance. Literature review showed that the evaluation of serious games is done in the field of education. A clear model specially developed for systematical evaluation of serious games in the field of education is MEEGA (Model for the Evaluation of Education Games) (Savi *et al.*, 2011). This model is explained in more detail in the next paragraph.

3.3 MEEGA(+) model

The model MEEGA (Model for the Evaluation of Education Games) (Savi *et al.*, 2011) is specially designed for the evaluation of educational games. The reaction of the participants with the use of a standardized questionnaire is captured. The model considers the reaction in terms of motivation, user experience, and learning. In addition, attention is given how to apply the evaluation model in practice. This evaluation approach proposed in the form of a case study (non-experimental). However, there is argued that the MEEGA model contains various limitations in its design and evaluation in terms of understanding of measurement instrument items and terms of validity and reliability. New research presents the MEEGA+ model, an advanced evaluation model of the MEEGA model. A factor analysis of the initial version of MEEGA model and literature review shows the overlap of the dimensions: motivation and user experience (Savi *et al.*, 2011). Therefore, the MEEGA+ model defined the *player experience* with factors including focused attention, fun, challenge, social interaction, confidence,

relevance, satisfaction, and usability (Petri *et al.*, 2016). The variable usability is subdivided into five dimensions: learnability, operability, aesthetics, accessibility, and user error protection (only used in digital games). An overview of this MEEGA+ framework is shown in figure 1.

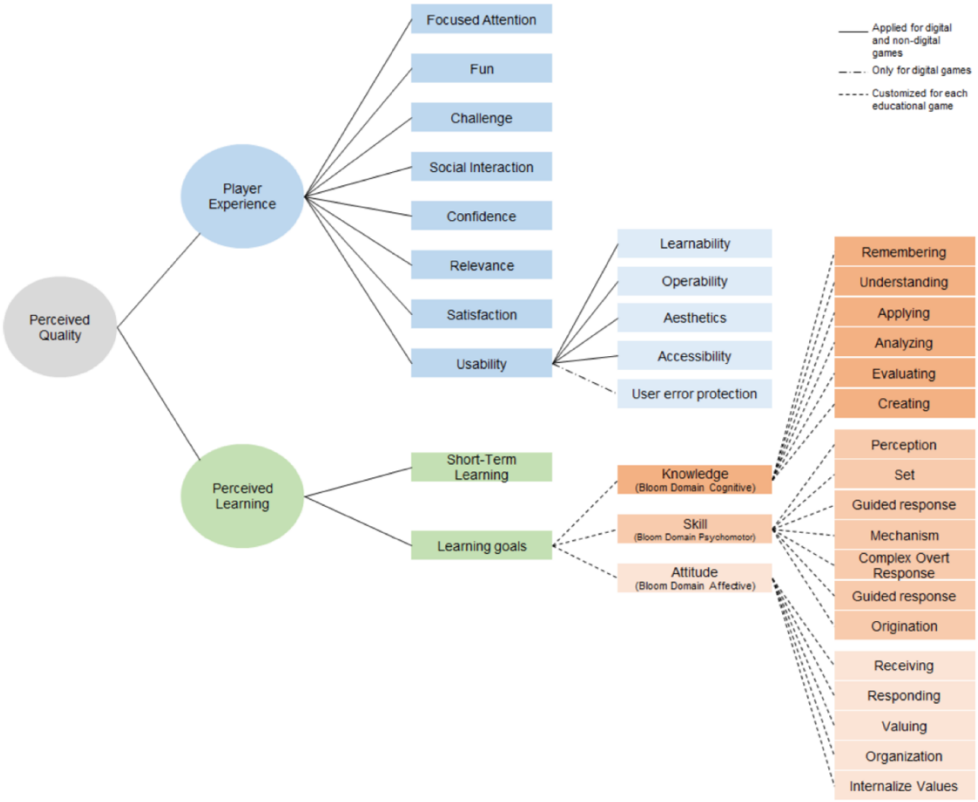


Figure 1 MEEGA+ framework

The variables focused on attention, confidence, relevance and satisfaction are not considered in this research as these variables are more relevant for the education field and are course related. For instance, relevance is measured by the item *It is clear to me how the contents of the game are related to the course.* Moreover, the variable usability is disregarded as this is less relevant for the content of the game. Usability is measured by: *The fonts (size and style) used in the game are easy to read.* The other variable's fun, challenge and social interaction measuring the player experience are considered as interesting to use as quality factors for serious board games in a business environment.

The evaluation of the learning goals with the model makes use of the revised version of Bloom's taxonomy of learning objectives is used. Bloom' three levels of learning are useful for the evaluation of knowledge at the cognitive level (remembering, understanding, and application), psychomotor domain (behavior and skills), affective domain (professional attitudes) (Petri *et al.*, 2016).

The potential benefit of the MEEGA+ model is the use of a standardized questionnaire whereby the perception of the participants is assessed. In this way, the evaluation can be completed in a relative non-instructive way and without much development. Additionally, the standardized questionnaire can increase the validity and reliability of data collected in comparison to self-assessments which can lead to low validity as data is collected via interviews or ad-hoc questionnaires. The MEEGA+ model stands out as a well systematically developed questionnaire since it explicitly decomposing evaluation goals into measures in the questionnaire. The evaluation focus on learning improvement, user experience during game play and motivation. Several studies from different authors used MEEGA model in practice with various games and context (Calderón, 2015).

Accordingly, the MEEGA+ model is considered as a promising framework for evaluating serious games. However, the connection to the use of serious board games in organizations is missing. Therefore, literature is used to find an accepted evaluation model in the business environment. The most popular evaluation model used in evaluating training programs is the four level model of Kirkpatrick. The next paragraph explains this model in more detail.

3.4. Kirkpatrick's' four level evaluation model

Donald L. Kirkpatrick published in 1959 the first ideas about measuring the effect of training programs. Kirkpatrick's model has now become one of the most known evaluation models (Oprins *et al*., 2015; Bates, 2004). Kirkpatrick's model is used to evaluate learning in organizations, including four levels of evaluation: reaction, learning, behavior, and results. These four levels with an explanation are shown in table 1.

Table 1 The four levels of Kirkpatrick's model (Oprins *et al*., 2015)

	Level	Definition
1	Reaction	Refers to the immediate reaction of the participants to the experience
2	Learning	Refers to learning effects, measured directly after the experience
3	Transfer	Refers to the transfer of learning in another context
4	Result	Refers to the impact on the targeted outcomes

Later, a fifth level was added by Phillips (1996) to calculate the *return on investment* (ROI). Moreover, Kaufman (1994) argued that the four levels are focused too much on the evaluation alone and missing evaluation of societal impact. Therefore, he suggested that another level of *societal impact* should be added to the model. This level considers the impact and consequences in and for society. This is related to Serious games covering sustainability issues. For instance, the CO_2 reduction issue is not only facing the industries but is also for social importance. Hence, measuring the social impact of these games is difficult.

The power of the model is that the four levels simplify the complex process of evaluating learning interventions (Oprins, 2015; Alliger *et al.*, 1989; Bates, 2004). The model simplifies the evaluation as it focusses on the data gathering after the training, which eliminates the measures of learning indicators before the training. Moreover, the surrounded factors influencing the training process are disregarded (Bates, 2004). Research shows that many have used the levels of Kirkpatrick's framework as a basis for their evaluation model (Hamblin, 1974; Molenda *et al.*, 1996; Tamkin *et al.*, 2002).

Although Kirkpatrick's framework is used often, the model has been much criticized. Three major critical points are pointed out and explained to what extent in this research is dealing with this disadvantage.

First, Salvator Falletta (1998) criticizes Kirkpatrick's model because it provides no instructions on how to decide if the model is applicable. For example, the model recommends using a control group and a pre- and post-test for measuring the behavior change and results, in levels three and four. Often, for several trainings, this is too much. Therefore, it is not needed to evaluate all steps of the model in all training, according to Salvatore Falletta (1998). A short training of a few hours deserves not as much evaluation as a week-long training. As in this research, there is a focus on evaluating games that approximately take one hour, the first three levels are considered.

Second, there is argue that the model is incomplete due to not considering individual or contextual influences (Bates, 2004). Bernthal (1995) discussed the model mixes evaluation and effectiveness. Kraiger and Jung (1997) suggest that Kirkpatrick provides *how* to evaluate model, and lacks *what* to evaluate and *how to link* the results to strategy. To have a complete evaluation, extra evaluation models are used in this research.

Lastly, in the model the assumption of causal linkages is made. For example, the model assumes that positive reactions lead to greater learning that delivers higher transfer and thus positive results. However, Kirkpatrick is still vague about what the precise nature is from the causal linkages. Most researchers failed to confirm such causal linkages. There is only little

evidence of the linear causality suggested by Kirkpatrick (1994) (Alliger *et al.,* 1997; Bates, 2004). In the conceptual framework of this research the linkages are not considered. However, analyses are done to investigate the linkages in this model.

The criticism that is stressed by different researchers indicate that Kirkpatrick's model cannot be used on it is own in evaluations. In addition, the model was designed for evaluating training programs and implementing it for evaluating serious games in teaching will ignore multiple quality characteristics (Abdellatif *et al.*, 2018). For that reason, other evaluation models are consulted that is applicable to the different levels of Kirkpatrick's and can be applied for serious games evaluation. Kirkpatrick's model is used as a basis to form an evaluation model. Each level is explained more in detail and complemented with the MEEGA+ and other models found in the literature.

3.4.1 Level 1: reaction

The first level of Kirkpatrick's model is *reaction*. This level measures the degree to which learners react to a serious game. This reaction level can entail several aspects, like motivation, usability or engagement (Steiner *et al.*, 2015). Analyzing the viewpoint of the participants may help to highlight the strengths and the weaknesses of the serious game.

As Kirkpatrick's model does not tell which variables need to be evaluated in this level, the MEEGA+ model is used to identify the elements influencing the perceived quality of the game. The variables fun, challenge and social interaction are considered as important and the reason for this is explained.

Fun is measured since it is essential to know if the participants have a positive attitude towards the serious game experience. To know the participants would like to play serious games another time. Also, a positive attitude supports learning (Boyle *et al.,*2011; Tamkin *et al.,*2002).

As stated by Pavlas (2010) and Susi *et al.* (2007) the game should have an appropriate level of *challenge*. Challenge refers to the actual content of the game, the problems the player is faced with. A balance between the challenge of the game and the skills of the players is needed (Waterman, 1990). When games are too easily solved they will not be engaged (Susi *et al.*, 2007). As skills generally improve during play, the challenge should increase as well.

There is a growing consensus among researchers about the positive effects of learning by stimulating people to discuss information and problem with different perspectives and to elaborate and refine them in order to reconstruct and co-construct (new) knowledge (Slavin, 1997). Therefore, in this research extra attention is on the social interaction during participants

in a serious game. Social interaction is divided into three levels: sharing, co-construction and constructive conflict as a higher level of interaction can increase the tacit knowledge sharing and in this case learning.

Nonetheless, in the MEEGA+ model little attention is given to the social interaction between the participants during the game. Considering that the integration of social interactions into serious games potentially improves learning and stimulates cooperation (Gee, 2003), more interest is in theory behind social interactions. The next paragraph gives more information about what social interaction is, levels of interaction and interaction activities.

3.4.1.1 Social interaction

Social interaction is about interpersonal (face-to-face) interaction and is assumed to be very important in games (Gee, 2003). In the game, it is often about playing together in cooperation or a competitive setting. Moreover, the reflective discussion in the debriefing session. Several studies show that the social context in which we learn can hugely influence how fast and how well we learn, and how motivated we are (Vygotsky, 1980; Reichart, 2014). Serious games are played in a team context. The advantage of playing in teams is that people can learn from the experiences of others, next to their own experience (Ickes *et al.*, 1994). Due to the interaction between the team members knowledge and skills can be transferred. This information exchange brings sources of knowledge together between team members (Ellis *et al*., 2003).

In literature, two types of knowledge are addressed: explicit and tacit knowledge (Polanyi, 1966; Nonaka, 1994; Haldin-Herrgard, 2000). Rivera-Vazquez *et al.* (2009) define knowledge sharing as the process of commonly exchanging tacit and explicit knowledge to create new knowledge.

Explicit knowledge is concrete, tangible, and identifiable data or information that is described in formal language. This knowledge is more directly observed, captures, transferred and communicated to others (Smith, 2001; Bennett, 1998). In contrast, tacit knowledge obtains personal knowledge that is gained by experiences, personal beliefs, and insights that is difficult to define. Therefore, it is difficult to manage, transfer or share this knowledge in everyday discussions and face-to-face meetings (Groff and Jones, 2012). Examples of tacit knowledge are the ability to ride a bicycle or the knowledge of an expert tennis player. An individual is usually reluctant to share his/her tacit knowledge with others (Osterloh & Frey, 2000) as tacit knowledge is seen as a skill or knowledge which gives possible advantages compared to others. Choi and Lee (2003) suggest that tacit knowledge of personal experiences can only be

exchanged with social interaction. Thus, a social relationship may be the most important factor that facilitates tacit knowledge sharing among people (Nonaka, 1994).

Conversion of tacit knowledge to explicit knowledge offers greater value to an organization (Haldin-Herrgard, 2000). Nonaka (1994) gives insight into how learning can result from mixing knowledge bases. The basic idea is that when people gather together in "communities of interaction", a dialogue exists between imlicit and explicit knowledge. The conversation held by people with a combination of tacit and explicit knowledge (experience and facts) can expand and build up understanding and can result in new knowledge. One type of knowledge stimulates the processing of the other (Bennett, 1998). This knowledge creation is considered as a process of interaction and negotiation with other participants in a learning environment (Kanselaar *et al.,* 2000). Serious games have the potency to facilitate interaction and support the opportunity of interaction and negotiation to makes the tacit knowledge explicit. It is from interest what types of social interactions are possible. The next paragraph goes deeper into the levels of interactions.

Levels of interaction

Decuyper *et al.* (2010) divide three levels of interaction: sharing, co-construction, and constructive conflict. *Sharing* refers to the process of communicating unshared knowledge, competencies, creative thoughts or opinions to other participants that were unknown before. *Co-construction* refers to combining insights and information by dialogues and reflective communication to share interpretations; and *constructive conflict*, refers to the process of dialogue, discussion, and negotiation that reveal different perspectives of participants. However, these three processes are very much interconnected and supportive. For instance, co-construction can lead to constructive conflict, in constructive conflict new information can be shared between participants. The social context in which these processes take place are from importance to reach shared cognition and thus collaborative learning (van den Bossche, 2006).

In conclusion, the first level of Kirkpatrick's model with the related quality elements is discussed. In the next paragraph, the second level of Kirkpatrick's model is explained in more detail.

3.4.2 Level 2: learning

Level two of Kirkpatrick's model is learning. Learning is a change in the learner's knowledge attributable to experience (Mayer, 2010). This level measures the learning that has taken place during the serious game and is measured right after playing the game. The evaluation will investigate to what degree the targeted knowledge is acquired by interacting with the serious game (Steiner *et al.,* 2015).

The way of evaluating learning is inspired by the MEEGA+ model. This model evaluates learning with the use of learning goals. The learning goals are made with the use of the Revised Bloom's Taxonomy (Anderson, 2006). The level of the learning goal can be classified in: cognitive, psychomotor or affective. *Cognitive learning goals* are used when the aim of the learning goal is to improve knowledge of understanding or remembering. As a game need to contribute to skill development, like problem-solving, communication or teamwork the learning goals can be classified in the *psychomotor domain* (Simpson, 1966). Learning goals related to growth in attitudes, emotions and feelings can be classified in the taxonomy of *affective domain* (Krathwohl *et al.*, 1973).

These levels give control to determine to what level the participants need to experience the learning goals in the game. Nevertheless, not only the level of learning is important in serious games but also the intention of the participant to use what they have learned in the game experience in another context. Subsequently, level three of Kirkpatrick's model is clarified.

3.4.3 Level 3: transfer

Kirkpatrick's third level of evaluation is transfer. This level addresses the degree to which learning is transferred in another context (e.g., work environment) (Oprins *et al.,*2015). Evaluation of transfer of learning directly in the real-world is most of the time impossible (Pined *et al.*, 2011). As it is costly based on human and economic resources. Moreover, it is questionable to what extent the game has contributed to this performance (Korteling *et al.*, 2011; Pinada *et al.*, 2011). Though, the game is ineffective if the participants are not able to use the material they have learned during the game, even if learning has taken place. However, the Kirkpatrick's model lacks on how to measure this level.

Therefore, other studies are reviewed here in literature. Some authors have suggested the need to find ways to measure transfer learning indirectly (Pineda *et al.,* 2011; Holton, 1997; Baldwin *et al.,* 1988; Pineda *et al.*, 2010). The study of Pineda *et al.* (2011) introduces a method wherein "intent to transfer" is indicated as a highly predictive variable of transfer.

In line with the Theory of Planned Behavior (Ajzens, 1991), "intent to transfer", could be a surrogate variable for transfer (Machin *et al.,* 2004). A valid and reliable model to determine transfer of the game is the FET (Factors for the Evaluation of Transfer) (Pineda *et al.*, 2011). However, this model does not provide the used items in the questionnaire. Based on these different theories, it is suggested that first there is need of a positive transfer intention to have transfer at all (Quesada-Pallarès *et al.,* 2015).

Further, level three of Kirkpatrick's model is followed up by level four, results. Level four is explained next.

3.4.4 Level 4. Results

The fourth level includes the measurement of the *results*, which implies the degree to which the targeted outcomes are a result of the game. This level is likely the most costly and time-consuming because the training needs to be linked to outcomes, benefits or final results and an effective way to measure this outcome in the long term needs to be created. Examples of results criteria include cost-savings, customer satisfaction or profitability (Alliger *et al.,* 1997). This level is outside the scope of this research. Since organizations choose to measure transfer instead of results, as results are difficult and the costs needed overcoming the difficulties may be too expensive (Cheng *et al.*, 2008). Moreover, the game is provided to different organizations and measuring the results is very specific for every organization. Consequently, it makes it hard to make a standard evaluation of the results.

3.5 Conceptual framework

Now that different variables have been introduced that have an influence on the quality of serious games, a conceptual framework is developed. The framework is based on level one, two and three of Kirkpatrick's model. The outcomes of the levels one and two are often easiest to measure and change. Therefore, these two levels provide the most useful information (Bernthal, 1995). However, level three, the transfer of learning into practice is also of interest. The three levels construct the dimensions: reaction, learning and transfer and are used to evaluate the quality of a serious game. The variables per level are illustrated in this chapter.

Level 1: Reaction

The first level of Kirkpatrick's model refers to this study to the thoughts of the participant about the game experience. The outcomes of reaction provide feedback that helps to evaluate learning experience and can be useful in the early stages of game development (Bee *et al.*, 2003). The data provides concrete quantitative data that can be used by decision-makers such as managers (Schumann *et al.*, 2014). In the framework, the sub-dimensions for the reaction are challenge, fun and social interaction. There is a growing consensus among researchers about the positive effects of learning by stimulating people to discuss information and problem with different perspectives and to elaborate and refine them in order to reconstruct and co-construct (new) knowledge (Slavin, 1997). Therefore, in this research extra attention is on the social interaction during participants in a serious game. Social interaction is divided into three levels: sharing, co-construction and constructive conflict as a higher level of interaction can increase the tacit knowledge sharing and in this case learning.

Level 2: Learning

Level two relates to the intended learning objectives and outcomes of the game in question. On this level, the evaluation will investigate whether and to what degree learners acquire the targeted knowledge by interacting with the serious game and the participants. Learning is the most evaluated since its presence converts the game into a serious game (Abdellatif *et al.*, 2018). The model assumes that games can contribute to the learning of technical knowledge and professional attitudes (Petri, 2017). The learning goals are used to examine if the participants experienced the dilemmas that are intended to come across during the game.

Level 3: Transfer

Transfer learning is the main goal that organizations look for when they invest in gaming (Pineda-Herrero *et al.,* 2014). As the evaluation of transfer directly is extremely difficult, different authors have little evidence that transfer can be measured indirectly and predict learning transfer. Therefore, the variable transfer intention is used in this model.

This chapter has illustrated how challenge, fun, social interaction, learning goals and transfer intention was expected to be evaluated to indicate if the serious game is from the perceived quality. The conceptual framework of the dependent variable (Perceived quality of the game) dependent the three levels of Kirkpatrick (reaction, learning and transfer) and on the independent variables (challenge, fun, social interaction, learning goals and transfer intention) is shown in figure 2. The perceived quality of the game is dependent on the reaction, learning, and transfer. The independent variables of reaction are challenge, attitude and social interaction. The independent variables of learning are the learning goals and the independent variables of transfer are the transfer intention.

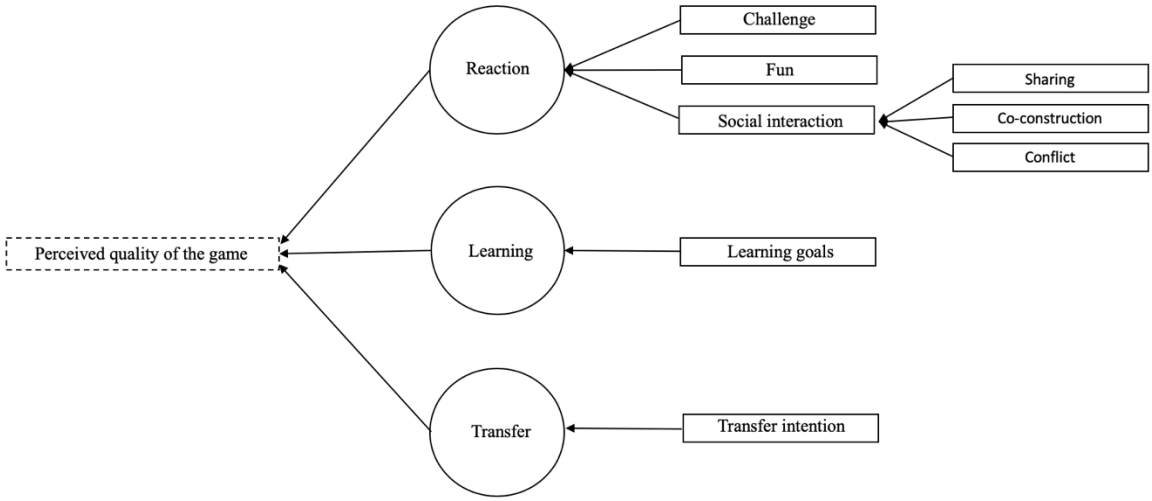


Figure 2 The dependent variable (Perceived quality of the game) is independent on the dependent variables (challenge, attitude, social interaction, learning goals and transfer intention)

To conclude, the objective of this thesis is to develop an evaluation model to systematically evaluate serious board games and explore which factors affect the perceived quality of serious board games in a business environment covering complex sustainability problems. By answering the question:

Which serious game elements contribute to the perceived quality of serious board games in a business environment to stimulate cooperation between different stakeholders in complex sustainability problems?

Sub-questions:

- Which existing evaluation models support the evaluation of serious board games?
- To what extent the evaluation model can be applied to different serious board games?
- In what way the various quality factors relate to each other?

4. Method

To measure the variables incorporated in the conceptual model a questionnaire is developed. There is chosen for a questionnaire as it is a non-disruptive way of evaluating. The questions can be filled in easy and quick by the participants, instead of doing an interview. Moreover, conducting the questionnaire is uniform. As there may be different facilitators the questionnaire can be hand out by every facilitator. In addition, the analysis is less time consuming when compared to analyzing video recordings. Furthermore, the data from the systematical evaluation questionnaire provide the researcher to compare different games as the same questions are asked.

The structure of the questionnaires is as follows. The three levels of Kirkpatrick are used as basis to develop the questionnaire and is complemented with the MEEGA+ model and other models from the literature. Two categories of items of the MEEGA+ model (reaction and learning) were omitted from the MEEGA+ questionnaire and expanded with the social interaction level questions from Van den Bossche *et al.* (2011). The questions used in the third level, transfer, are inspired by the questionnaire designed by Anthony Machin and Gerard Fogarty (2004) measuring transfer intention.

The general study design is a one-shot post-test only design. The questionnaire contains three parts with 34 items and five open questions. These open questions are used to conduct qualitative data. Through a questionnaire the evaluation goal is assessed based on the three levels of Kirkpatrick's model. This research design can be performed without lots of extra effort in a relative non-disturbing way when compared with experiments with control groups or qualitative interviews (Petri *et al*., 2016).

Different criteria guided the selection of the instruments. First, the MEEGA+ model is validated one-shot post-test only design, easy to customize to the different objectives of serious games and already applied to games. However, this model focusses on the education field. To connect this model to the business environment, Kirkpatrick's model is used. As social interaction is important to increase the cooperation and negotiation between the stakeholder's extra attention is given to the levels of social interaction. The MEEGA+ model lacks on the evaluation of the third level of Kirkpatrick's model. Therefore the model of Anthony Machin and Gerard Fogarty (2004) has been accessed.

The response format of the questionnaire is a 5-point Likert scale with response alternatives ranging from strongly disagree to strongly agree based on the MEEGA+ model (Petri *et al.*, 2016). The 5-point format allows the respondent to express the opinion about the

items with precision. Besides, a neutral point is added allowing being comfortable to express their opinion and thereby contributing to the quality of the answers (Petri *et al.,* 2016*)*. In the next paragraphs, the different sections of the questionnaire are described.

4.1 Measures

4.1.1Reaction

The first section of the instrument evaluated reaction addressing three concepts: challenge, fun, social interaction. The reaction is measured using the MEEGA+ model on the dimensions: challenge, fun and social interaction (Petri, 2018). The following items are measured:

Challenge

Items concerned the level of challenge of the game showed a low level of internal consistency ($\alpha = 0.65$). Since the items are validated by Petri *et al.,* (2018) and the variable considered as important, the items were computed into one mean variable.

- The game provides new challenges offers new obstacles, situations or variations at an appropriate pace
- This game is appropriately challenging for me

Fun

The attitude of the participant to the serious games experience is measured by the following items (Petri *et al.,* 2018; Von Wangeheim *et al., 2012).* These items were computed into one mean variable as it showed an acceptable level of internal consistency ($\alpha = 0.72$).

- I would play a serious game again
- I would recommend this game to my colleagues/friends

Social interaction

Social interaction is measures by items of Petri *et al.,* (2018). Participants rated the items on a scale from strongly disagree (1) to strongly agree (5). The items were computed into one mean variable as it showed a high level of internal consistency ($\alpha = 0.88$)

- I felt good interacting with other players during the game
- I was able to interact with other players during the game

The idea to stimulate social interactions between the participants with the game leads to focus on the construction of the aspects: sharing, co-construction and constructive conflict. These aspects were measured by means of nine items based on the questionnaire of Van den Bossche *et al.* (2011).

Sharing

The sharing of information that was before unshared is measured by the following items. The level of internal consistency is low ($\alpha = .58$). Since the used items are validated in previous studies the items are used for analysis.

- During the game, I shared all relevant information and ideas I had
- Participants were listening carefully to each other

Co-construction

The process of co-construction starts when one of the team members describes the problem situation whereas the other team members are listening. This shared information is refined and changed in some way due to questioning, concretizing and completing the information. The shared knowledge gets new meaning that was not available to the team yet (Webb *et al.*, 1996). Co-construction is measured by means of the following items ($\alpha = 0.70$):

- Participants elaborate on each other's information and ideas
- Information from participants is complemented with information from others

Constructive conflict

Team members discuss different perspectives with each other through negotiation and dialogues. During this process different opinions are uncovered and conflict can occur which leads to further communication. This is done in an open and productive manner, meaning that they are approachable to opinions of other team members and are willing to reach an agreement. Disagreement in a team can give more insight into different ideas and perspectives. This level of interaction is measured by the following items ($\alpha = 0.79$):

- This team tends to handle differences of opinions by addressing them directly
- Comments on ideas are acted upon
- Opinions and ideas of participants were verified by asking each other critical questions

Besides asking questions related to the social interaction, also observations are done to verify the activities. Three levels of activities can be identified: cognitive, affective and metacognitive. The player activities with the subcategories are shown in table 2.
The scheme is based on the coding scheme developed by Else Veldhuis-Diermanse (2002).

Table 2 Observing coding scheme of players activity related to learning activities

Learning activity	Players activity
Cognitive learning activities	Debating
	Using external experience and information
	Linking or repeating internal information
Affective learning activities	General
	Asking for feedback
	Chatting/social talk
Metacognitive learning activities	Planning
	Keeping clarity
	Monitoring

4.1.2 Learning

Inspired by the MEEGA+ model, learning goals measure the level of learning. Each learning goal should be customized in the questionnaire with use of the following statement: The game contributed to <*verb related to the level of the learning goals*> < goal>. For example, in accordance with the learning goals of ZeroBrine Game, a game to increase the circular economy of brine, a statement would be: "The game contributed to *understand* the need of circular resource recovery". After the game session the level of agreement is captured. Participants rate the items on a scale from strongly disagree (1) to strongly agree (5). Just like Petri *et al.* (2016), these ratings are used to define the contribution to learning of the game.

4.1.3 Transfer

Inspired by the questionnaire designed by Anthony Machin and Gerard Fogarty (2004) and Petri *et al.* (2016), the intention to transfer knowledge of the participant is measured. As stated by Rouiller and Goldstein (1993), thinking about opportunities of using the knowledge learned is a significant unique variance in transfer. The statements covering this variance are used in the questionnaire. Machin and Fogarty have a focus on improving skills, however in serious games the focus is on increasing knowledge. Therefore, the statements are reformed

in the context of serious games. These items were computed into one mean variable as it showed an acceptable level of internal consistency (α = 0.79).

- After playing the game, I am motivated to start a discussion/dialogue with other participants
- I will definitely spend time thinking about how to use the knowledge that I have gained

Moreover, as every game supports specific transfer of knowledge there is a possibility to modify the evaluation. Therefore, statements can be customized. This is possible by adapting the next statement: After playing the game, I will <*verb related to the behavior change*> <intention behavior change>. For example: "After playing the game I will change my attitude of sharing data with other parties"

4.1.4 Control variables

In addition, the characteristics of the sample are identified. The years of work experience is asked (0 - 1 years = 1, 2 - 3 years = 2, 4 - 6 years = 3, 7 - 10 years = 4, > 10 years = 5). The years of work experience can have influence on the effect of learning and the intention of behavior change. There is asked in which institute they are working (1 = business and industry, 2 = civil society, 3 = NGO's, 4 = policy, 5 = science, 6 = student) and what was the role in the game (1 = government, 2 = industry, 3 = market, 4 = technology provider, 5 = knowledge institute, 6 = transport, 7 = distributor, 8 = producer). From this information, there can be measured if the game is more efficient if participants have the same or a different role than in real life. There is measured if the participant is working in the food branch (yes = 1, no = 0), and if so, is the game topic relevant for the food industry? (yes = 1, why? , no = 0). This information is relevant to make a connection with the food industry. Moreover, there is from interest if people playing the game together know each other already (yes, I knew everyone = 2; I knew some of them = 1; No, I did not know anybody = 0), considering that this variable can influence the level of interaction. To have an overview of the overall reaction of the participants to the game experience the grade is asked (1 = very poor, 10 = excellent). Moreover, this grade can be useful for the marketing of the game.

4.2 Sample and procedure

A pre-test for the questionnaire was conducted to make sure the survey did not contain any mistakes. Three fellow students were asked to check the questionnaire on clearness of the items. Moreover, this version was tested at the WE-Energy game to get insights on how participants would react to the questionnaire. In addition, the questionnaire is checked in collaboration with Renate Wesselink (Education and Learning Science at WUR). Her experience in doing research in learning and development of professionals in both companies and schools helped to improve the questionnaire. Additionally, to meet the requirements of the mandatory the questionnaire is orally discussed with Frans van den Akker (ISPT).

Data collection took place in March and April of 2019 in Rotterdam and Amersfoort. At four different game moments the participants were asked to fill out the questionnaire after playing the serious game. The different games are described below.

ZeroBrine game

ZeroBrine game is a serious game designed to create awareness on the scarcity of water and brines. The aim of the game is to support social interaction and connect different stakeholders. The game consists of 5 rounds. In each round, the water availability decreases. The industry, government, mineral market and technology providers experiencing the dilemmas and consequences of decisions that are made. ISPT developed the game. This game moment is used for two data points. Between the data points the game is optimized. The first data point is called version one (v1) and the second data point is called version two (v2).

CO_2 reduction game

CO_2 reduction game is a serious game that is designed to create awareness of the consequences of CO_2 emission and the complexity of reducing the CO_2 emission. The game consists of seven rounds. In each round the industries, government and technology providers need to take decisions. The industries are producing CO_2 emission and can reduce these emissions by buying technologies at the technology providers. However, these technologies are expensive and it takes a while before the technologies can be implemented. The government can influence the policies in the country, collect CO_2 taxes and has subsidies to help the industries to innovate. The game was developed in the commission of ISPT.

Hydrogen game

The hydrogen game is a serious game that is designed to give insights into a hydrogen transition for transport. Hydrogen could be a possible solution for CO_2 emission reduction. However, there are still several dilemmas around the hydrogen transition. In the game these dilemmas are experienced to create awareness on the possibilities and challenges of hydrogen use in the transport sector. ISPT developed the game together with Perspectivity.

Participants were recruited via the network of ISPT with the use of e-mail and phone. Participants were invited to play the game. The game takes around one hour. Afterwards, a collective discussion is taken place with the participants and the facilitator. Right after this, the participants were asked to fill in the questionnaire to collect the data. Data collection took place during four game moments in March and April of 2019.

The participants (n = 51) participating at the games in this study were men and women from different nationalities, but mainly Dutch. The respondents in this study had different work field backgrounds ranging from business and industry (8.9 %), civil society (0.0 %), NGO's (6.7 %), policy (4.4 %), science (4.4 %), student (75.6 %) or other (0.0 %). Except the students, the years of work experience is asked at the other participants (n = 13) differ from 0 – 1 year (8.9 %), 2 – 3 years (6.7 %), 4 – 6 years (6.7 %), 7 – 10 years (2.2 %) or > 10 years (4.4 %).

As there is interest in the application of these games in the food industry, the respondents were asked if they are working in the food sector (26.7 %) and if so, there is asked for their opinion if they see a link between the food sector and the subject of the game (92.3 %).

4.3 Data analysis

After the data collection with google forms, the data was exported to Excel. From Excel, the data was exported to IBM SPSS Statistics (version 24) to carry out data analysis. Afterwards, the data have been checked for outliers and incomplete responses. The three games had different numbers of learning goals analyzed. The least relevant learning goals were removed from the analysis as this brought some impossibilities for analyzing.
A reliability analysis was conducted to confirm scales. All four data points (game moments) are used for reliability analysis for the dimension reaction and transfer. Items were considered reliable with a Cronbach's alpha $\alpha > 0.7$ (Santos, 1999). One item was excluded before computing items into one mean variables as the Cronbach's alpha increased significantly. For challenge, the item 'The game does not become monotonous as it progresses (repetitive or boring tasks)' assuming that the remaining items of still measure the level of challenge of the game. This was the only item that has a negative formulated statement. After having items computes into one mean variable, the relationship between all relevant variables has been explored and described using correlation analysis.

Multiple linear regression analysis was done, as with this analysis the effect of multiple predictor variables (sharing, co-construction and conflict) on the dependent variable social interaction could be indicated. Another multiple linear regression analysis was conducted to check the effect of the variables challenge, fun and social interaction on reaction. As there is a possibility that the variables in the model measuring the same information a multicollinearity test was conducted (Hair *et al.,* 2019). Kirkpatrick's model does not provide the order of the levels. Therefore, a hierarchical linear regression was performed to identify the effect of the various variables on the level learning and transfer (table 3). Effect sizes were calculated for Models 2, 3, and 4 using Cohen's f^2 (Cohen, 1988)

Table 3 Four variations of the hierarchical regression analysis

Dependent variable	Variables			
	Model 1	Model 2	Model 3	Model 4
Transfer	Reaction	Learning	Challenge, fun, social interaction	Sharing, co-construction, conflict
Transfer	Learning	Reaction	Challenge, fun, social interaction	Sharing, co-construction, conflict
Learning	Reaction	Transfer	Challenge, fun, social interaction	Sharing, co-construction, conflict
Learning	Transfer	Reaction	Challenge, fun, social interaction	Sharing, co-construction, conflict

To check the overall agreement on the variable with the use of the 5-point Likert-scale, descriptive statistics were conducted. For the analysis of items of each variable, the quality of the game referred to three classifications (Rajab *et al.,* 2011). The mean score and the corresponding interpretation are shown in table 4. The variable that has a mean score from 1 to 2.33 response to the poor quality level. Meanwhile, if the mean score is between 2.34 and 3.67 the quality level of the game reaches the average or neutral level. Additionally, if the mean score is between 3.68 and 5 it is interpreted that the variable has a good level of quality.

Table 4 Interpretation of Mean score

Mean score	Interpretation
< 2.33	Poor
2.34 - 3.67	Moderate
> 3.67	Good

5. Results

This section presents the most relevant findings.

5.1 Informal observation of the participants while playing the serious game

The participants seemed to be very enthusiastic about the idea of playing a serious game, which adds to the widespread perception that games are motivational **(**Bourgonjon *et al.*, 2011). During the game session, the participants were deeply engaged and immersed in the game as they mention many times that the time flued by. In the beginning at every game there is still confusion how to play the game. Therefore, there is agreed with Pavlas (2010) and Susi *et al*. (2007) that it is important that the game is not too challenging. There is seen that after one round the players are more familiar with the game rules and an increase in dialogues, negotiation and strategic discussions on the subject of the game starts.

5.2 Evaluation serious board game experiences

In order to test the quality of the serious games, the mean of the eight variables are conducted by making up the questionnaire (see Appendix I) after playing the game. All items in the questionnaire are rated at a 5-point Likert-scale (strongly disagree = 1, strongly agree = 5). In order to compare the different games on the variables in this study, the mean per variable was conducted and classified in levels (table 4). The higher the value of the mean suggesting that the variable was more experienced in the game by the participants. First the variables related to the reaction are explained, followed by learning and lastly the Means of the intention to transfer are shown.

5.2.1 Reaction

For the overall reaction to the game the participants were asked to give the game experience an overall grade from one to ten. Hereby, the one indicates a poor experience and a ten an excellent experience. The results are shown in table 5.

Table 5 Overview Mean of general grade to the game experience

		General grade	
	N	**Mean**	**SD**
ZeroBrine version 1	12	7.75**	.87
ZeroBrine version 2	6	7.67**	.52
Hydrogen	27	6.74**	1.20
CO2 reduction	6	7.83**	.98

* Significant at p < 0.05 (2-tailed); ** Significant at p < 0.01 (2-tailed)

The given grade for the hydrogen has the highest standard deviation (SD = 1.20), indicating that the overall reaction of the participants is highest divided. Moreover, this game has also the lowest mean (M = 6.74). The CO_2 reduction game is highest rated with a mean of M = 7.83.

For the dimension reaction the variable challenge, fun and social interaction were measured. The mean score, standard deviation and the quality level of these variables in the game are shown in table 6 & 7.

Table 6 Overview mean for variable challenge and fun

		Challenge			Fun		
	N	**Mean**	**SD**	**Level**	**Mean**	**SD**	**Level**
ZeroBrine v1	12	3.83**	.91	Good	4.42**	.87	Good
ZeroBrine v2	6	3.92**	.49	Good	4.33**	.60	Good
Hydrogen	27	3.50**	.73	Moderate	3.93**	.69	Good
CO2 reduction	6	4.00**	.45	Good	4.67**	.52	Good

* Significant at p < 0.05 (2-tailed); ** Significant at p < 0.01 (2-tailed)

The analysis of the challenge showed that the mean score of the CO_2 reduction game has the highest mean (M = 4.00) and the hydrogen game has the lowest mean (M = 2.85). Observations of the game moment of the hydrogen game gave the impression that the game was too challenging. This was also verified by feedback of the participants naming: *"There is need of a better introduction at the beginning, and gradually increase complexity. Now it is too complex at the start", "Each role should be explained clearer and more supervision of the instructor is needed".* The CO_2 received feedback as: *"Give more background information". "Explain more the technologies". "Give more explanation what is happening in one round".* Looking to two versions of the ZeroBrine game, the mean of version one to version two of the increases from M = 3.83 to M = 3.92. Indicating that the version two had a better challenge level than version one.

The attitude of the participant to the game experience was evaluated with the variable fun. In overall, the participants rated the experience positive as the Mean is around 4 or higher, shown in table 6. Participants mention positive feedback on the attitude of the game: *"fun to play and learned a lot in an interactive way", "informative", "teamwork with other players, realistic and fun".*

The variable social interaction measures the possibilities to interact with other participants during the game (table 7). Both versions of the ZeroBrine game have the highest mean (M = 4.25) and the hydrogen the lowest (M = 3.81). Overall the social interaction is high rated and is confirmed with the feedback from the participants: *"Good interaction", "Good start for discussions", "Start discussions about the interest of each stakeholder", "Discussion in a multidisciplinary setting". "Good teamwork needed between everybody, it gives good insight in the interest of different people and organizations".*

Table 7 Overview means for variable social interaction and sharing

		Social interaction			Sharing		
	N	**Mean**	**SD**	**Level**	**Mean**	**SD**	**Level**
ZeroBrine v1	12	4.25**	.89	Good	4.04**	.89	Good
ZeroBrine v2	6	4.25**	.41	Good	3.67**	.41	Good
Hydrogen	27	3.81**	.89	Good	3.26**	.89	Moderate
CO2 reduction	6	4.16**	.75	Good	3.17**	.75	Moderate

* Significant at $p < 0.05$ (2-tailed); ** Significant at $p < 0.01$ (2-tailed)

The sharing of information decreases from version one to version two of the ZeroBrine game. Moreover, the hydrogen and CO_2 reduction game score moderate, respectively. The degree of co-construction of information and the level of conflict is measured, shown in table 8. The hydrogen scores moderated on this variable, whereas the three other games have better scores. Observing the hydrogen game, there was absolutely high interaction. However, more affective and metacognitive activities were observed.

Table 8 Variable mean for variables co-construction and conflict

		Co-construction			Conflict		
	N	**Mean**	**SD**	**Level**	**Mean**	**SD**	**Level**
ZeroBrine v1	12	4.08**	.90	Good	4.13**	.74	Good
ZeroBrine v2	6	3.75**	.27	Good	4.33**	.26	Good
Hydrogen	27	3.54**	.78	Moderate	3.65**	.55	Moderate
CO2 reduction	6	3.83**	.41	Good	4.09**	.49	Good

* Significant at p < 0.05 (2-tailed); ** Significant at p < 0.01 (2-tailed)

Next to the reaction of the participants the level of learning goals experienced in the game are from importance. Therefore, participant opinions were asked to what extend they experienced a certain learning goal. The results are pointed out in the next paragraph.

5.2.2 Learning goals

The level of learning goals experienced by the participants in the game are evaluated. The results of agreement on the learning goals incorporated into the game are visible in table 9.

Table 9 Variable mean for variable learning goals

		Learning goals		
	N	**Mean**	**SD**	**Level**
ZeroBrine v1	12	4.20**	.54	Good
ZeroBrine v2	6	4.23**	.15	Good
Hydrogen	27	3.69**	.73	Good
CO2 reduction	6	4.40**	.31	Good

Both ZeroBrine games and the CO_2 game have high agreement (M = 4.20; M = 4.23; M = 4.40) on experiencing the learning goals. Participants said about the CO_2 game: *"Gives good insights in the need for* CO_2 *emission reduction".* The hydrogen has the lowest score (M = 3.69), still participants mentioned positive statements: *" I learning about energy flow", "Shows the important role of the government".*

Besides the measurement of what they have learned, there is interest in the participants' intention to use in another context what is learned in the game experience. Therefore, the results of transfer intention per game is explained in the next paragraph.

5.2.3 Transfer intention

The intention to transfer the knowledge gained during the game is measured. The mean and the level per game of this variable is shown in table 10.

Table 10 Overview mean for variable transfer intention

		Transfer intention		
	N	**Mean**	**SD**	**Level**
ZeroBrine v1	12	4.00**	.93	Good
ZeroBrine v2	6	3.92 **	.20	Good
Hydrogen	27	3.60**	.87	Moderate
CO2 reduction	6	4.33**	.61	Good

* Significant at p < 0.05 (2-tailed); ** Significant at p < 0.01 (2-tailed)

The transfer intention after playing the CO2 reduction game is highest (M = 4.33). This game scores also highest on the other variables, there is a possibility that if the higher the other variables are rated this have influence on the transfer intention.

In this chapter only attention is given to the variables on its own. There is also curiosity how the correlation is between the variables, as the conceptual model assumes that some of the variables are correlation to each other. In the next paragraph, the correlations between the variables are expressed.

5.3 Correlation matrix

The relationship between the variables of the perceived quality of serious games was investigated using Pearson product-moment correlation coefficient to investigate how the variables are related to each other. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and homoscedasticity. A correlation matrix was conducted (table 11). There is a strong, positive correlation between all the variables observed, N = 51, $p < .05$. The VIF values were less than three, which indicate on low multicollinearity (Hair *et al.,* 2019).

Table 11 Correlations among variables

	Variable	Correlations							
		1	2	3	4	5	6	7	8
1.	Reaction								
2.	Challenge	.70**							
3.	Fun	.51**	.52**						
4.	Social interaction	.41**	.33*	.52**					
5.	Learning	.76**	.63**	.68**	.62**				
6.	Sharing	.64**	.56**	.59**	.64**	.64**			
7.	Co-construction	.45**	.57**	.62**	.46**	.60**	.61**		
8.	Conflict	.59**	.45**	.53**	.45**	.69**	.58**	.53**	
9.	Transfer	.63**	.59**	.61**	.43**	.73**	.55**	.67**	.54**

N = 51

* Significant at $p < 0.05$ (2-tailed); ** Significant at $p < 0.01$ (2-tailed)

5.4 Multiple regression analysis

In the conceptual framework there is assumed that as well as, social interaction and reaction are predicted by different variables. A linear regression analysis was conducted to test the effect of the predictor variables (sharing, co-construction and social interaction) on social interaction and the effect of fun, challenge and social interaction on the reaction. First, the effect on social interaction will be discussed, followed by showing the correlation and regression of the variables (challenge, fun and social interaction) to reaction.

5.4.1 Social interaction

The effect of the variables sharing, co-construction and conflict on social interaction on the dependent variable is checked by use of correlation and linear regression. The correlation between sharing, co-construction and conflict on social interaction is conducted by correlation matrix, shown in figure 4. The variable sharing has the highest correlation (r = .70) with social interaction and co-construction and conflict respectively .46 and .45.

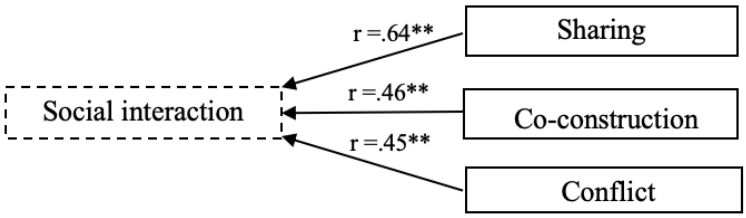


Figure 3 Correlation matrix of social interaction with sharing, co-construction and conflict

The regression analysis revealed a significant value for sharing on social interaction, suggesting that the more sharing of information and ideas has taken place, the more social interaction is experienced (table 12). The linear regression analysis with the predictor variable sharing and the dependent variable social interaction showed a regression coefficient of 0.47 (t (3) = 3.607; p = 0.001). The total variance in the dependent variable *social interaction* is by 15.8% uniquely explained by the variable *sharing*.

Table 12 Results of Linear Regression Analysis

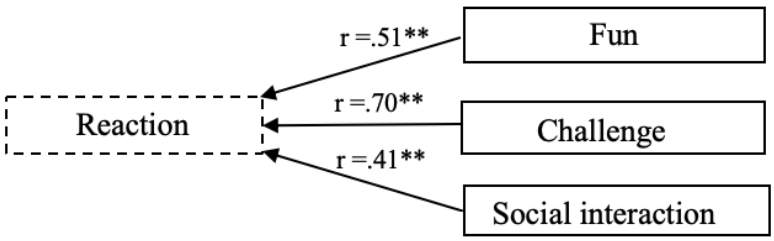
	Social interaction					
	b	**SE B**	**β**	**Sig.**	**Part (%)**	**VIF**
(Constant)	1.64	0.56		0.005		
Sharing	0.47	0.13	0.54	0.001	15.84	1.85
Co-construction	0.09	0.15	0.09	0.533	0.48	1.71
Conflict	0.10	0.18	0.08	0.562	0.43	1.64
Df	3					
F	11.71**					
Adjusted R2	0.39					

N = 51

* Significant at p < 0.05 (2-tailed); ** Significant at p < 0.01 (2-tailed)

5.4.2. Reaction

The effect of the predictor variables challenge, fun and social interaction on the variable reaction is checked by use of correlation and linear regression. The variable challenge has the highest correlation (r = .70) with reaction and social interaction the lowest (r = .41).



* Significant at p < 0.05 (2-tailed); ** Significant at p < 0.01 (2-tailed)

Figure 4 Correlation between reaction and fun, challenge and social interaction

The results of table 13 show a significant value for *challenge*, suggesting the more an appropriately level of challenge is in the game, the higher the overall grade to the game experience is given. The linear regression analysis, with the predictor variable challenge and the dependent variable *reaction*, showed a regression coefficient of 0.89 (t (3) = 4.993; p = 0.000). The total variance in the dependent variable *reaction* is by 25% uniquely explained by the variable *challenge*.

Table 13 Results of Linear Regression Analysis

	Reaction					
	B	**SE B**	***β***	**Sig.**	**Part (%)**	**VIF**
(Constant)	2.21	0.76		0.005		
Challenge	0.89	0.18	0.58	0.000	23.8	1.38
Fun	0.19	0.20	0.13	0.331	1.0	1.68
Social interaction	0.23	0.17	0.15	0.201	1.7	1.38
Df	3					
F	17.97**					
Adjusted R2	0.50					
N = 51						

* Significant at p < 0.05 (2-tailed); ** Significant at p < 0.01 (2-tailed)

The previous analysis done where related to parts of the conceptual model. In order to get more insight in the relations of the variables in the complete model, hierarchical regression analysis is done. The next paragraph, shows the results of four hierarchical regression analysis. In two analysis transfer is used as dependent analysis and the other two analysis learning is used as dependent variable. Variations are made with the variables using in Model 1 and Model 2 to identify the model which order of variables fits best the conceptual model.

5.5 Hierarchical regression

A disadvantage of Kirkpatrick's model is that there is no relation between the different evaluation levels given. Therefore, a hierarchical regression analysis was conducted to test the relations of all predictor variables on the dependent variables learning and transfer. First, two hierarchical linear regression analysis were done to check the predictive variance of sharing, co-construction and conflict, after controlling for fun, challenge and social interaction, learning and reaction on transfer. Next, two hierarchical linear regression analysis were conducted to check the predictive variance of the variables on learning.

5.5.1. Transfer as dependent variable

Table 14 summarizes the results of the regression analysis with transfer as dependent variable and reaction in Model 1 and learning in Model 2. Learning (ß = .40 , p < .05) and co-construction (ß = .34) were positively correlated with transfer, indicating the more learning and co-construction was experienced in the game the intention of transfer increases.

Table 14 Hierarchical regression with reaction, learning, challenge, fun, social interaction, sharing, co-construction and conflict as predictors of transfer

D.V.	Transfer			
	Model 1	Model 2	Model 3	Model 4
IVs				
Reaction	.63**	.16	.09	.16
Learning		.61**	.47*	.40*
Challenge			.14	.02
Fun			.20	.10
Social interaction			-.05	-.07
Sharing				-.03
Co-construction				.34*
Conflict				-.02
N	51	51	51	51
R2	0.39	0.55	0.58	0.64
Adjusted R2	0.38	0.53	0.54	0.57
ΔR2		0.16	0.04	0.05
*f*2	31.57**	16.35**	1.28	2.02

Standardized coefficients.
* Significant at p < 0.05 (2-tailed); ** Significant at p < 0.01 (2-tailed)

Adding learning to the model yielded a significant effect on transfer (f^2= 16.35; adjusted R^2 = .53). Furthermore, adding challenge, fun and social interaction on the model yielded not significant effect on transfer (f^2 = 1.28; adjusted R^2 = 0.54). Sharing, co-construction and conflict adding to the model give also no significant effect on transfer (f^2 = 2.02; adjusted R^2 = .57). Indicating that learning accounts for the highest variance on transfer intention.

As there are doubts if learning or reaction should be used in Model 1, also a hierarchical analysis is done including model 1 the variable learning and Model 2 the variable reaction. Table 14 summarizes the results of the regression analysis. However, still learning (ß = 0.40, $p < .05$) and co-construction (ß = .34, $p < .000$) show a positive correlation with transfer.

Table 15 Hierarchical regression analysis with learning, reaction, challenge, fun, social interaction, sharing, co-construction and conflict as predictors of transfer

D.V.	Transfer			
	Model 1	Model 2	Model 3	Model 4
IVs				
Learning	.73**	.61**	.47**	.40*
Reaction		.16	.09	.16
Challenge			.14	.02
Fun			.20	.10
Social interaction			-.05	-.07
Sharing				-.03
Co-construction				.34**
Conflict				-.20
N	51	51	51	51
R^2	0.54	0.55	0.58	0.64
Adjusted R^2	0.53	0.53	0.54	0.57
ΔR^2		0.01	0.04	0.05
f^2	56.46**	1.16	1.28	2.02

Standardized coefficients.
* Significant at $p < 0.05$ (2-tailed); ** Significant at $p < 0.01$ (2-tailed)

As there is also interest in the variance of the variables in the evaluation model to learning, the hierarchical regression with dependent variable learning is conducted.

5.5.2. Learning as dependent variable

The results of the regression analysis with learning as dependent variable are shown in table 15. The results showed significant positive correlation effects between reaction (ß = .35 . p < .05), transfer (ß = .22 , p < .05), social interaction (ß = .26, p < .05) and conflict (ß = .20, p < .05) with learning. Suggesting that more learning occurs if the variables are experienced.

The results of the regression analysis are shown in table 15 Adding reaction ((f^2= 68.12; adjusted R^2 = .57), transfer (f^2=16.35; adjusted R^2 = .68), challenge, fun and social interaction ((f^2= 6.02; adjusted R^2 = .75) to the model yielded a significant effect on learning. However, adding sharing, co-construction and conflict to the model does not give any significant effect. Showing that these three variables have less influence on learning in the game.

Table 16 Hierarchical regression analysis with reaction, transfer, challenge, fun, social interaction, sharing, co-construction and conflict as predictors of learning

D.V.	Learning			
	Model 1	Model 2	Model 3	Model 4
IVs				
Reaction	.76**	.50**	.38	.35*
Transfer		.42**	.25	.22*
Challenge			.04	.07
Fun			.18	.16
Social interaction			.24	.26*
Sharing				-.13
Co-construction				.01
Conflict				.20*
N	51	51	51	51
R^2	0.58	0.69	.78	.80
Adjusted R^2	0.57	0.68	.75	.76
ΔR^2		0.11	.09	.03
f^2	68.12*	16.35*	6.02*	1.75

Standardized coefficients.
* Significant at p < 0.05 (2-tailed); ** Significant at p < 0.01 (2-tailed)

Since learning can also be explained by the variable transfer, a hierarchical analysis is conducted with transfer in Model 1 and in Model 2 the variable reaction. The results of this analysis are shown in table 17 .

The analysis showed significant positive correlation effects between transfer (ß = .22, p < .05), reaction (ß = .35, p < .05), social interaction (ß = .26, p < .05) and conflict (ß = .20, p < .05) with learning. Suggesting that the more these variables are experienced in the game the more learning occur.

Table 17 Hierarchical regression analysis with transfer, reaction, challenge, fun, social interaction, sharing, co-construction and conflict as predictors of learning

D.V.	Learning			
	Model 1	Model 2	Model 3	Model 4
Transfer	.73**	.42**	.25*	.22*
Reaction		.50**	.38*	.35*
Challenge			.04	.07
Fun			.18*	.16
Social interaction			.24	.26*
Sharing				-.13
Co-construction				.01
Conflict				.20*
N	51	51	51	51
R^2	0.54	0.69	0.78	.80
Adjusted R^2	0.53	0.68	0.75	.76
ΔR^2		0.15	0.09	.03
f^2	56.46**	23.51**	6.02*	1.75

Standardized coefficients.
* Significant at p < 0.05 (2-tailed); ** Significant at p < 0.01 (2-tailed)

The analysis does not show other correlation effects in comparison to the analysis with reaction in Model 1. However, the results show the variance of transfer in Model 1 and reaction in Model 2 the model is more significant (f^2= 56.46; p < .000). Indicating that this order of variables in the model could be promising.

6. Conclusion and discussion

This section contains the interpretation and explanation of the data. Outcomes are interpreted within the context of the research questions that were posted at the beginning of this thesis. Overall, this study aimed to give insights into the factors influencing the perceived quality of serious board games and provide a systematical evaluation framework to evaluate serious games. Participants were asked to fill out a survey after playing a serious board game with a sustainability topic used in a business environment. The questionnaire contains different questions regarding the quality of the game.

This thesis had identified some game characteristics in an attempt to identify the quality of a serious board game used in a business environment. The strengths of the evaluation model is that there is a balance between broad application scope through flexibility and standardized questions to compare different serious board games easily. Moreover, this design forced the game designer to define clear learning goals from the beginning and to stick to it during the game development process.

The findings tell us that an evaluation model needs to contain different quality factors that can be addressed in three levels of evaluation: reaction, learning, and transfer. First, the reaction of the participant after playing the game need to be evaluated. This is captured by the quality factors of the positive attitude after playing the game, the level of challenge of the game, and the social interaction experienced during the game. Testing the evaluation model on four game sessions shows that the use of this evaluation model reaches its objectives in terms of getting insights into the game experience of the participants according to challenge, fun, social interaction, learning, and transfer.

Overall, observing the game sessions the game experience was experienced as positive. This is confirmed with the agreement on variable fun and the overall grade the participants filled in on the questionnaire. There is indicated that challenge is the best predictor for the reaction of the game. Other studies also state that the challenge is an important quality factor (Petri *et al.,* 2016; Cowley *et al.*, 2014). This is probably because it is important that the player at least can understand the game in such a way that the player is able to participate. As this is not possible, there can be expected that the game experience is affected significant negatively. This was also observed at the hydrogen game. One participant did not understand the rules fast enough and quite the game. Moreover, the hydrogen game had the lowest rating on the challenge of the game. Observing the game moment, the game was too challenging through incompleteness of the explanation and facilitating skills of the moderator.

Further, the interaction activities were more on affective and metacognitive level. The ZeroBrine game version one to version two increases in agreement on the challenge variable. As in version one, the feedback came: “Make the game harder, for instance, make it more complicated to trade between companies”. Therefore, in version two the complexity of working together between the companies was increased, which probably had an influence on the level of challenge observed.

The findings did not support the expectation that a higher level of social interaction would influence the interaction between the participants. The variable sharing had the highest predictor factor for social interaction. A reason for this could be that only a minimal number of players participating at the game. Consequently, there was more interaction between players with different roles as no full teams could be formed. The level of sharing of information was lower at the second version of the ZeroBrine game compared to version one. This could be explained by the change in game design. In version one the participants were sitting all at the same table, as in version two the participants were divided through the room having their own table. In this way, more assertiveness was needed to share knowledge with the other participants. The hydrogen has the lowest mean on the variables, this was unexpected when observing the game session as there was a lot of interaction between the participants. However, as the challenge was high the interactions were more based on the game on itself and not about the content. The hydrogen game was played in a room with three game tables at the same time, which made it harder for the facilitator to have the lead over the group.

The second level of evaluation is learning. The findings suggest that the extent of learning in the serious game is influenced by the quality factors reaction, transfer, social interaction and conflict. The learning goals in the ZeroBrine game and CO2 game are best incorporated, suggesting the highest mean in table 8.

The third level of evaluation is transfer and refers to the intention to use what they have learning within the experience in another context. It was identified that transfer is for the most part explained by the experiences on learning and co-construction. Suggesting that learning and co-construction have a positive influence on the intention of transfer.

Overall, it can be assumed that in the case of the serious games, the model of evaluating the quality of the game is best promising with learning as dependent variable. The predictive variables are in the order of transfer, reaction, challenge, fun and social interaction. The three levels of interaction (sharing, co-construction and conflict) do not give significant variance for explaining learning. However, due to a small respondent number (n = 51) no prove of changing

the model can be given. Therefore, in this research the quality of the game is dependent from the balanced combination of the eight variables considered.

Still, the questionnaire has possibilities for improvements. After observing the game sessions, it stands out that the role of the facilitator is from significant importance. Responding to the open feedback questions the recurrent adjective is the clearness and completeness of the explanation at the beginning of the game. In order to help the facilitator improving the game explanation there is recommended to add a closed question that covers what the participant thinks about the completeness of the explanation at the start of the game. With an option for an open opinion what they would recommend changing in the game explanation.

To identify what level of challenge is important for which category of players and the facilitator can ask at the start of the game if the participants have already experienced serious games. A more challenging version can be played. The items asked on challenge give limited information. Recommend to add items like: “The game was difficult to understand and play”, “It was really difficult to understand the final goal or objective of the game” (Alamri *et al.,* 2014).

The number of learning goals included in the questionnaires of the different games were different. For instance, the hydrogen game had more items covering learning goals than the ZeroBrine game. This gave some inconvenience in analyzing the data. Therefore, there is recommended to always use the same number of learning goals when comparing different games. Therefore, in this method there is chosen to formulate the five most important learning goals to include in the questionnaire.

Additionally, there is recommended to add more items per variable as an evaluation only two items per variable gave inconvenience with data analysis.

There was a difference observed in time it takes of understanding the game between participants that are more familiar with playing serious games than others for which it was the first time. Therefore, there is recommended to determine their prerequisites beforehand and take them into account in the analysis and interpretation of the results (Emmerich *et al.,* 2016).

Concerning these limitations of the questionnaire, an adapted questionnaire is developed and presented in Appendix III .

The serious game framework shows the major components that create an effective model for learning and transfer through the use of serious board games. There is demonstrated that serious games will not succeed just because they are games with educational content. All the component inside this framework plays a role to ensure that learning would take place while playing the game. This framework formed as an appropriate basis for effective serious games design for designers. The serious games used in this paper offer a promising tool in the communication of complex sustainability challenges with different stakeholders in business environments.

6.1 Limitations

A limitation of this study is that a small number of game sessions with a low number of participants were done. Executing simulation games is time-consuming for both, the participants and the researchers, therefore it makes it more difficult to find participants to join the game. As a result, it is hard to get a large amount of data statistically significant conclusions. The results were more sensitive for outliers and can be less used for generalization.

Additionally, as there are only three or less items analyzed per variable there is a reasonable change that the variables were not embrace the whole construct. Thus, conclusions could have been done on variables that are not completely measured.

Moreover, this study was conducted mainly among students, however, when conducting this study with experts' different results could be observed due to differences in the knowledge on the topics and more into the relations of the stakeholders in business life (Emmerich *et al.*, 2016). Students understand the game play earlier than the older players which possibility influenced the challenge variable positively.

In addition, the games use for data gathering were all in development. This could cause an effect on the evaluation of the game, as the organization or facilitation could be improved. The more often the game is facilitated this could be limited and this could improve the quality of the game.

Measuring a reaction after a game session is not without complications. Participants may be incapable of adequately evaluating the value of that experience. Moreover, one aspect of the serious game experience may influence other aspects of the game, even though they may be separate dimensions. For example, dissatisfaction with the facilities may have a negative effect on the respondent's rating on the learning experience (Schumann *et al.*, 2014).

6.2 Further research

Future research is needed to address questions that remained unanswered by this research. First of all, while doing the research there is found that the quality characteristic of transfer intention after playing a serious board game is limited analyzed (Abdellatif *et al.*, 2018).
Besides, the fourth level of Kirkpatrick's model to evaluate the results of serious game play could be investigated. However, there is a need to consider the viability of this, because the game play is only around one hour and evaluating the level four, results, takes a lot of investments economically and to evaluate this level. Moreover, as the results want to be measured in the business environment different types of organizations need to take into consideration.

7. Recommendation

As evaluation does not finish at the point that data is collected, but the data should lead to a process of revision to make the game more effective and appealing. Therefore, a framework in combination with a spreadsheet is conducted to make the revision of the data more easily. In this chapter, an evaluation-driven design framework is provided (Emmerich *et al.*, 2016) and explanation of the spreadsheet inspired by the MEEGA+ model is given (Petri *et al.,* 2016).

The framework of evaluation-driven design offers guidance to the role of evaluation during the design and development process. The framework consists of three phases: the preparation phase, the design phase and the evaluation phase. The design and evaluation phase are interlinked with each other. The framework is intended to model the whole process. On each process a lot of research has been done, but will not formulated in detail as each element is a broad topic on itself.

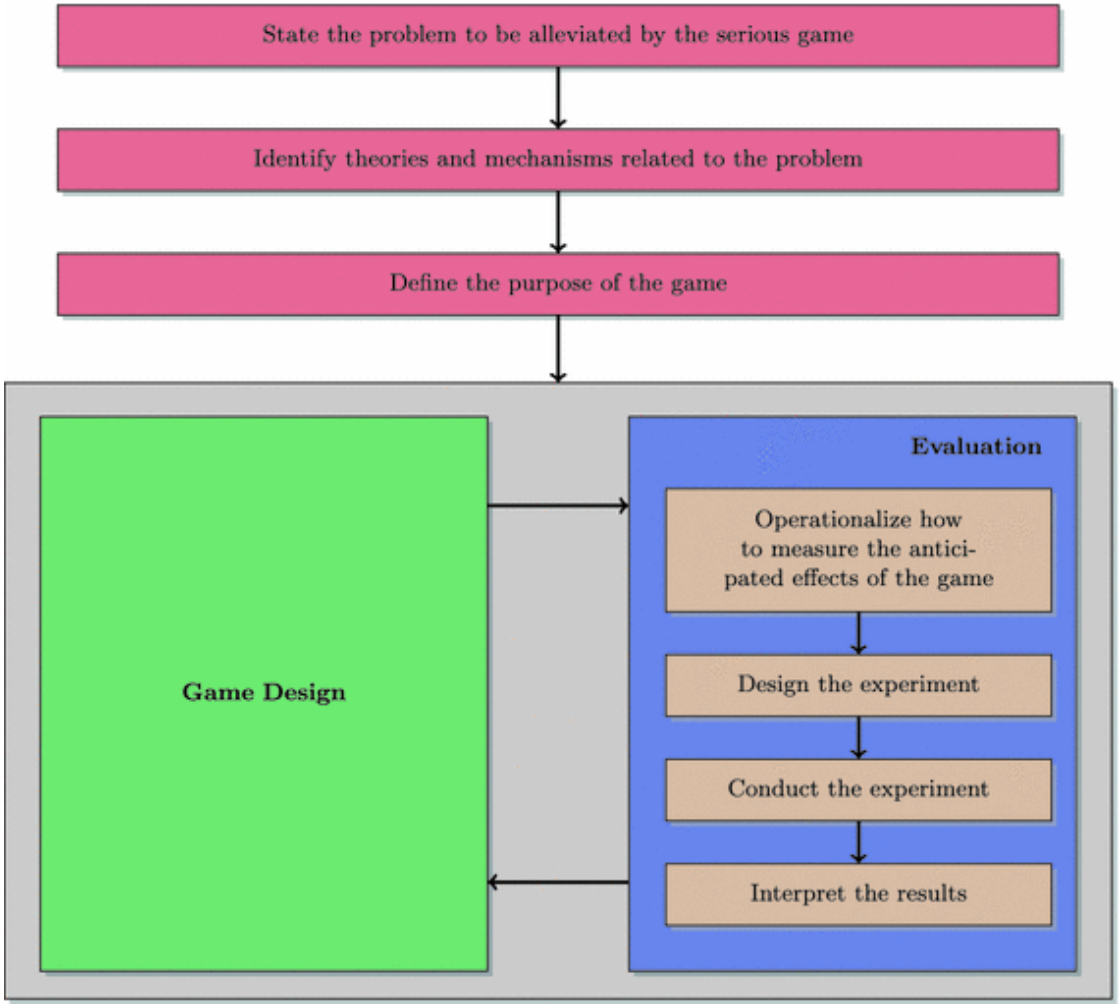


Figure 5 Framework *of Evaluation-Driven Design for serious games (Emmerich et al., 2016)*

Every serious game project starts with describing the *problem statement*. As the game need to contribute to improvement of the situation, the following up step is to *identify the theory behind the problem*. The factors and underlying processes need to be defined which is needed to translate the problem into a game. Afterwards, based factors influencing the problem, the *definition of the purpose of the game* can be made. The purpose is later used to evaluate the game (Emmerich *et al.,* 2016). With help of the GQM goal template (Basili *et al.,* 1994) the evaluation objective can be defined: Analyze the <*name of the selected game*> for the purpose of <*intention of the evaluation*> in terms of <*quality aspects*> from the < *from who's point of view*> in the context of <*environment of the context* >.

The game design and evaluation process are tightly linked to each other. In order to design the game in a gradually way, a prototype can be used for testing. The results of this test can be used for the improvement of further design.

The evaluation phase consists of three main phases. In the first step the *desired effect* of the game needs to be analyzed, here for the purpose of the game needs to be clear. In this step the data collection instrument, provided in this thesis need to be customized for the evaluation of the selected game. The information related to the evaluation context can be inserted in advance in the questionnaire: game's name, place and date (figure 9).

GAME TITLE: Hydrogen game								
	Dimensions							
	Subdimensions		Fun		Challenge		Social interaction	
	Key word		Recommend	Play again	Challenging	New challenges	Able to interact	Good interaction
	Item nr in survey		1	2	3	4	5	6
		1	-1	1	-1	0	-2	0
		2	1	1	-2	0	0	1
		3	0	1	1	1	2	1
		4	1	1	-1	0	-1	1
		5	-1	1	-1	0	1	1
		6	1	1	-1	1	1	1
		7	1	1	0	0	1	-1
		8	1	1	0	1	2	1
		9	1	1	0	1	1	1
		10	1	2	1	1	1	1
		11	0	0	-2	-1	-1	-1
		12	1	1	1	0	2	0
	Date: 15-03-2019 / Rotterdam Haven	13	1	1	0	1	0	1
		14	2	1	0	1	1	1
		15	2	1	1	1	1	0
		16	-1	1	-1	-1	1	1
		17	1	1	0	1	1	1

Figure 6 Game's name, place and date game in spreadsheet

Afterwards, the learning goals should be customized to the learning objectives of the game. Choose the five most important learning objectives of the game. Therefore, the following statement should be adapted: The game contributed to <*verb related to the level of the learning goal (cognitive, psychomotor, and affective)*> <goal/concept>. Figure 10 demonstrates an example how to present the learning goals of hydrogen game. It is also possible to add other statements for other quality factors. However, since this is outside of the standard questionnaire it is not possible to assure the validity and reliability.

Learning goals				
Acceptation, adaptation, implementation	Push/pull market	Cost structure	Need for scalability/innovation	Progress and consequences
16	17	18	19	20
0	-1	1	-2	-2
1	1	1	1	1
1	1	1	1	1
1	0	1	1	0
0	1	0	1	0

Figure 7 Learning goals

Organize a pilot as experiment to test whether the game has the intended impact on the player. The questionnaire is provided at the end of the game, right after the game discussion with the facilitator and participants. The questionnaires can be conducted manually, the questionnaire should be distributed to the participants by paper. However, the data collection is also possible through an electronic form of the questionnaire using Google Forms. The form can be provided by QR-code. The results should be interpreted carefully. The results give input for the design process to adapt changes.

7.1 Analyzing the results by spreadsheet

1. Customize the spreadsheet

Each row of the spreadsheet represents an answer of a participant. A column represents the items of the questionnaire and are categorized per level of evaluation and variable.
The researcher needs to adjust the spreadsheet by filling in the game title, date and place. Moreover, the learning objectives need to be customized. Figure 11 shows the place in the spreadsheet where to add the learning goals.

Learning goal 1	Item learning goal 1
Learning goal 2	Item learning goal 2
Learning goal 3	Item learning goal 3
Learning goal 4	Item learning goal 4
Learning goal 5	Item learning goal 5

Figure 8 Overview learning goals fill-in

2. Prepare collected data for analysis

The evaluations conducted manually must be prepared for the data analysis to fill them in the spreadsheet provided by this thesis. The data gathered via the electronic questionnaire can

be exported to excel directly. This data needs to be filled in or exported in the first sheet: data. The data is linked to the other sheets.

3. Demographics

The demographic information is presented in graphs in the Demographic graphs sheet (Figure 12)

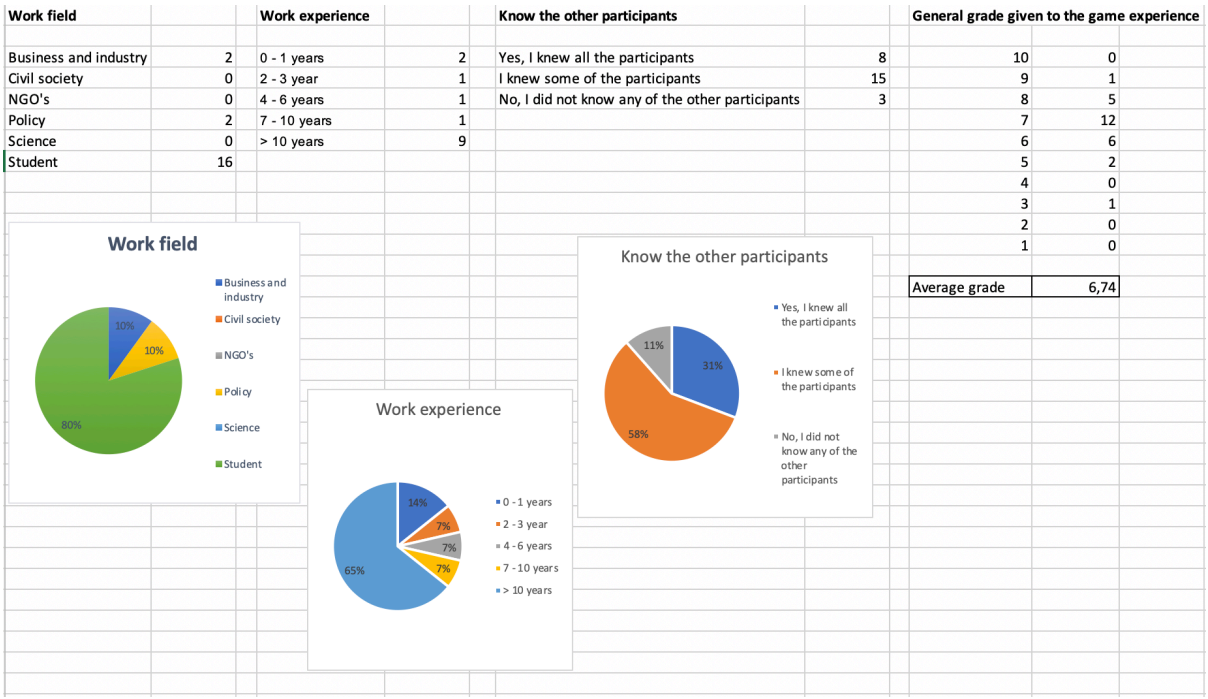


Figure 9 Overview demographics sheet

4. Descriptive statistics

From the data sheet the descriptive statistics are automatically calculated. The spreadsheet shows the frequency of response for each measurement instrument item, as presented in figure 13.

	Tot.	27	27	27	27	27	27	27	27	27
Frequency of responses	2	2	2	1	1	5	0	2	3	1
	1	13	18	8	15	14	15	16	4	19
	0	6	4	8	8	4	6	6	3	7
	-1	5	3	7	3	3	5	3	12	0
	-2	1	0	3	0	1	1	0	5	0
Reference to matrix	2	2	2	2	2	2	2	2	2	2
	1	1	1	1	1	1	1	1	1	1
	0	0	0	0	0	0	0	0	0	0
	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
Percentage of responses	2	7,4%	7,4%	3,7%	3,7%	18,5%	0,0%	7,4%	11,1%	3,7%
	1	48,1%	66,7%	29,6%	55,6%	51,9%	55,6%	59,3%	14,8%	70,4%
	0	22,2%	14,8%	29,6%	29,6%	14,8%	22,2%	22,2%	11,1%	25,9%
	-1	18,5%	11,1%	25,9%	11,1%	11,1%	18,5%	11,1%	44,4%	0,0%
	-2	3,7%	0,0%	11,1%	0,0%	3,7%	3,7%	0,0%	18,5%	0,0%
	Tot.	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Figure 10 Frequency of reponses per item

From this data, graphics of frequencies are automatically generated and presented in the "Graphs" sheet (Figure 14). The order of the graphics is based on the quality factors and dimensions of the questionnaire.

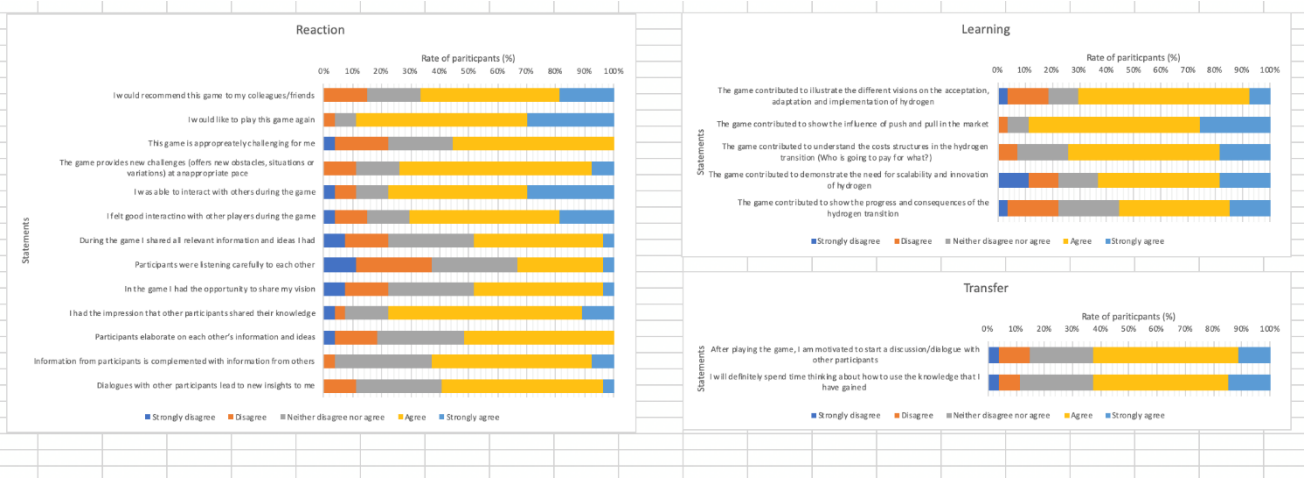


Figure 11 Graphs frequencies of responses per item

5. Discuss the results

The findings of the evaluation results, indicates the main contribution of this game but also the improvement opportunities. These results can be used to compare different game moments or games with each other.

References

Abdellatif, A. J., McCollum, B., & McMullan, P. (2018, March). Serious games: Quality characteristics evaluation framework and case study. In *2018 IEEE Integrated STEM Education Conference (ISEC)*, 112-119.

Abt, C. C. (1987). *Serious games*. University press of America.

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.

Alamri, A., Hassan, M. M., Hossain, M. A., Al-Qurishi, M., Aldukhayyil, Y., & Hossain, M. S. (2014). Evaluating the impact of a cloud-based serious game on obese people. *Computers in Human Behavior, 30,* 468-475.

Alliger, G.M. & Janak, E.A. (1989) 'Kirkpatrick's levels of training criteria: thirty years later', *Personnel Psychology, 42,* 331–342.

Anderson, L. (2006, May). Revised Bloom's taxonomy. Paper presented at North Carolina Career and Technical Education Curriculum Development Training, Raleigh, NC.

Baldwin, C. J. (2015). *The 10 principles of food industry sustainability*. John Wiley & Sons.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. Personel Psychology, *41*(1), 63-105.

Basili, V. R. (1994). Goal question metric paradigm. *Encyclopedia of software engineering*, 528-532.

Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and program planning*, *27*(3), 341-347.

Bennett III, R. H. (1998). The importance of tacit knowledge in strategic deliberations and decisions. *Management Decision*, *36*(9), 589-597.

Bernhaupt, R. (Ed.). (2010). *Evaluating user experience in games: Concepts and methods*. Springer Science & Business Media.

Bernthal, P. R. (1995), 'Evaluation that goes the distance', *Training and Development*, *49*(9), 41-45

Bourgonjon, J., Valcke, M., Soetaert, R., De Wever, B., & Schellens, T. (2011). Parental acceptance of digital game-based learning. *Computers & Education*, *57*(1), 1434-1444.

Bouwmans, M., Runhaar, P., Wesselink, R., & Mulder, M. (2017). Fostering teachers' team learning: An interplay between transformational leadership and participative decision-making?. *Teaching and Teacher Education*, *65*, 71-80.

Boyle, L., Hancock, F., Seeney, M., & Allen, L. (2009). The implementation of team based assessment in serious games. *In 2009 Conference in Games and Virtual Worlds for Serious Applications*, 28-35.

Bremmers, H. J. (2004). Dynamics in chains and networks: proceedings of the sixth international conference on chain and network management in agribusiness and the food industry. *Wageningen Academic Pub*.

Brown, K. G., & Gerhardt, M. W. (2002). Formative evaluation: An integrative practice model and case study. *Personnel Psychology*, *55*(4), 951-983.

Bruinsma, J. (2009). The Resource Outlook to 2050: By how much do land, water and crop yields need to increase by 2050? Paper presented at the FAO Expert Meeting, 24–26

Calderón, A., & Ruiz, M. (2015). A systematic literature review on serious games evaluation: An application to software project management. *Computers & Education*, *87*, 396-422.

Cheng, E. W., & Hampson, I. (2008). Transfer of training: A review and new insights. *International Journal of Management Reviews*, *10*(4), 327-341.

Chin, J., Dukes, R., & Gamson, W. (2009) Assessment in simulation and gaming: a review of the last 40 years. *Simul. Gaming 40*(4), 553–568

Choi, B., & Lee, H. (2003). An empirical investigation of KM styles and their effect on corporate performance. *Information & Management*, *40*(5), 403-417.

Connolly, T., Stansfield, M., & Hainey, T. (2009). Towards the development of a games-based learning evaluation framework. In *Games-based learning advancements for multi-sensory human computer interfaces: Techniques and effective practices*, 251-27.

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. *Comput. Educ. 59*(2), 661–686.

Cowley, B., Fantato, M., Jennett, C., Ruskov, M., & Ravaja, N. (2014). Learning when serious: Psychophysiological evaluation of a technology-enhanced learning game. *Journal of Educational Technology & Society*, *17*(1), 3-16.

Decuyper, S., Dochy, F., & Van den Bossche, P. (2010). Grasping the dynamic complexity of team learning: An integrative model for effective team learning in organisations. *Educational Research Review, 5*(2), 111-133.

DeSmet, A., van Ryckeghem, D., Compernolle, S., Baranowski, T., Thompson, D., Crombez,

G., Poels, K., van Lippevelde, W., Bastiaensens, S., van Cleemput, K., Vandebosch, H., de Bourdeaudhuij, I. (2014). A meta-analysis of serious digital games for healthy lifestyle promotion. *Prev. Med. 69*, 95–107

Dieleman, H., & Huisingh, D. (2006). Games by which to learn and teach about sustainable development: exploring the relevance of games and experiential learning for sustainability. *Journal of Cleaner Production*, *14*(9-11), 837-847.

Donald, K., & James, D. K. (1994). Evaluating training programs: four levels. *Anonymous Berrett Koehler San Francisco*.

El-Nasr, M. S., Drachen, A., & Canossa, A. (2016). *Game analytics*. Springer London Limited.

Ellis, A. P. J., Hollenbeck, J. R., Ilgen, D. R., West, B. J., Porter, C. O. L. H., & Moon, H. (2003). Team learning: Collectively connecting the dots. *Journal of Applied Psychology*, *88*, 821-835.

Emmerich, K., & Bockholt, M. (2016). Serious games evaluation: processes, models, and concepts. In *Entertainment Computing and Serious Games*, 265-283.

Falletta, S. V. (1998). Book Review: Evaluating Training Programs: The Four Levels. *American Journal of Evaluation*, *19*(2), 259-261.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, *1*(1), 20-20.

Girard, C., Ecalle, J., Magnan, A. (2013). Serious games as new educational tools: how effective, are they? a meta-analysis of recent studies. *J. Comput. Assist. Learn*.,29(3), 207–219

Göbel, S., Gutjahr, M., & Steinmetz, R. (2011, October). What makes a good serious game conceptual approach towards a metadata format for the description and evaluation of serious games. In *5th European Conference on Games Based Learning*, 202-210.

Groff, T., & Jones, T. (2012). *Introduction to knowledge management*. Routledge.

Hainey, T., Connolly, T. M., Stansfield, M., & Boyle, E. A. (2011). Evaluation of a game to teach requirements collection and analysis in software engineering at tertiary education level. *Computers & Education*, *56*(1), 21-35.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2-24.

Haldin-Herrgard, T. (2000). Difficulties in diffusion of tacit knowledge in organizations. *Journal of Intellectual capital*, *1*(4), 357-365.

Hamblin A C (1974), *Evaluation and Control of Training*, McGraw Hill

Haug, C., Huitema, D., & Wenzler, I. (2011). Learning through games? Evaluating the learning effect of a policy exercise on European climate policy. *Technological Forecasting and Social Change*, *78*(6), 968-981.

Hofstede, G. J., De Caluwé, L., & Peters, V. (2010). Why simulation games work-in search of the active substance: A synthesis. *Simulation & Gaming*, *41*(6), 824-843.

Holton III, E. F., Bates, R. A., Seyler, D. L., & Carvalho, M. B. (1997). Toward construct validation of a transfer climate instrument. *Human Resource Development Quarterly*, *8*(2), 95-113.

Ickes, W., & Gonzalez, R. (1994). "Social" cognition and social cognition: From the subjective the to the intersubjective. *Small Group Research, 25*, 294-315.

ISPT. (2019). *A trust based network - ISPT*. [online] Available at: https://www.ispt.eu/a trust-based-network/ [Accessed 23 Jan. 2019].

Juhola, S., Driscoll, P., de Suarez, J. M., & Suarez, P. (2013). Social strategy games in communicating trade-offs between mitigation and adaptation in cities. *Urban Climate*, *4*, 102-116.

Kanselaar, G., Andriessen, J., de Jong, T., & Goodyear, P. (2000). New technologies. In *New learning*, 55-81.

Kaufman, R., & Keller, J. M. (1994). Levels of evaluation: beyond Kirkpatrick. *Human Resource Development Quarterly*, *5*(4), 371-380.

Kirkpatrick, D. I. (1998). Evaluating Training Programs: The four levels. 2nd Edition

Kirkpatrick, J. (2009). The Kirkpatrick four levels: A fresh look after 50 years [white paper].

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation*. Association for Talent Development.

Klabbers, J. H. (1989). On the improvement of competence. In *Simulation-Gaming*, 3-7.

Korteling, J. E., Helsdingen, A. S., Sluimer, R. R., Van Emmerik, M. L., & Kappé, B. (2011). *Transfer of Gaming: Transfer of training in serious gaming*.

Kraiger, K., & Jung, K. M. (1997). Linking training objectives to evaluation criteria. *Training for a rapidly changing workplace: Applications of psychological research*, 151-175.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1973). Taxonomy of educational objectives, the classification of educational goals. Handbook II: affective domain. David McKay Co. *Inc., New York*, *1*, 956.

Krebs, C. (2003). The antibiotic battle. *Agri Marketing*, *41*(3), 34-34.

Kriz, W. C. (2003). Creating effective learning environments and learning organizations through gaming simulation design. *Simulation & Gaming*, *34*(4), 495-511.

Leemkuil, H., & De Jong, T. O. N. (2012). Adaptive advice in learning with a computer based knowledge management simulation game. *Academy of Management Learning & Education*, *11*(4), 653-665.

Lieberman, D. A. (2006). What can we learn from playing interactive games. *Playing video games: Motives, responses, and consequences*, 379-397.

Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious games analytics: Theoretical framework. In *Serious games analytics,* 3-29.

Machin, M. A., & Fogarty, G. J. (2004). Assessing the antecedents of transfer intentions in a training context. *International Journal of Training and Development*, *8*(3), 222-236.

Maloni, M. J., & Brown, M. E. (2006). Corporate social responsibility in the supply chain: an application in the food industry. *Journal of business ethics*, *68*(1), 35-52.

Mayer, I. (2012). Towards a comprehensive methodology for the research and evaluation of serious games. *Procedia Computer Science*, *15*, 233-247.

Mayer, I., Bekebrede, G., Harteveld, C., Warmelink, H., Zhou, Q., van Ruijven, T., & Wenzler, I. (2014). The research and evaluation of serious games: Toward a comprehensive methodology. *British Journal of Educational Technology*, *45*(3), 502 527.

Mayer, R. E. (2010). Applying the science of learning to medical education. *Medical education*, *44*(6), 543-549.

Medema, W., Furber, A., Adamowski, J., Zhou, Q., & Mayer, I. (2016). Exploring the potential impact of serious games on social learning and stakeholder collaborations for transboundary watershed management of the St. Lawrence River Basin. *Water*, *8*(5), 175.

Michael, D. R., & Chen, S. L. (2005). *Serious games: Games that educate, train, and inform*.

Molenda M, Pershing J, & Reigeluth, C. (1996), 'Designing instructional systems', *The ASTD Training and Development Handbook: a Guide to Human Resource Development*

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization science*, *5*(1), 14-37.

Notarnicola, B., Hayashi, K., Curran, M. A., & Huisingh, D. (2012). Progress in working towards a more sustainable agri-food industry. *Journal of Cleaner Production*, *28*, 1-8.

O'Neil, H.F., Wainess, R., & Baker, E.L. (2005). Classification of learning outcomes: evidence from the computer games literature. *Curriculum J. 16*(4), 455–474

Oprins, E., Visschedijk, G., Roozeboom, M. B., Dankbaar, M., Trooster, W., & Schuit, S. C. (2015). The game-based learning evaluation model (GEM): measuring the effectiveness of serious games using a standardised method. *International journal of technology enhanced learning*, *7*(4), 326-345.

Osterloh, M., & Frey, B. S. (2000). Motivation, knowledge transfer, and organizational forms. *Organization science*, *11*(5), 538-550.

Pavlas, D. (2010). A Model Of Flow And Play In Game-based Learning The Impact Of Game Characteristics, Player Traits, And Player States.

Peklaj, C. (2006). Cooperative activity and its potential for learning in tertiary education. *Horizons of Psychology*, *15*(3), 37-50.

Petri, G., von Wangenheim, C. G., & Borgatto, A. F. (2016). MEEGA+: an evolution of a model for the evaluation of educational games. *INCoD/GQS*, *3*.

Petri, G., & von Wangenheim, C. G. (2017). How games for computing education are evaluated? A systematic literature review. *Computers & education*, *107*, 68-90.

Petri, G., von Wangenheim, C. G., & Borgatto, A. F. (2018). MEEGA+: A Method for the Evaluation of Educational Games for Computing Education.

Phillips, J. J. (1996). ROI: The search for best practices. *Training & Development*, *50*(2), 42-48.

Pineda, P. (2010). Evaluation of training in organisations: a proposal for an integrated model. *Journal of European Industrial Training*, *34*(7), 673-693.

Pineda-Herrero, P., Quesada, C., & Stoian, A. (2011). Evaluating the efficacy of e-learning in Spain: a diagnosis of learning transfer factors affecting e-learning. *Procedia-Social and Behavioral Sciences*, *30*, 2199-2203.

Pineda-Herrero, P., Quesada-Pallarès, C., & Ciraso-Calí, A. (2014). Evaluation of training transfer factors: The FET model. In *Transfer of learning in organizations*, 121-144.

Polanyi, M. (1966). The logic of tacit inference. *Philosophy*, *41*(155), 1-18.

Quesada-Pallarès, C., & Gegenfurtner, A. (2015). Toward a unified model of motivation for training transfer: A phase perspective. In *Motivationsforschung im Weiterbildungskontext*, 107-121

Cohen, J. (1988). Statistical power analysis for the social sciences.

Rajab, A., Rahman, H. A., & Shaari, R. (2011). The international students perception towards

the education quality. *International Journal of Social Sciences and Humanity Studies*, *3*(2), 49-58.

Rana, P., Platts, J., & Gregory, M. (2009). Exploration of corporate social responsibility (CSR) in multinational companies within the food industry. *Queen's Discussion Paper Series on Corporate Responsibility Research*, *2*, 2009.

Reichart, B., & Bruegge, B. (2014, July). Social interaction patterns for learning in serious games. In *Proceedings of the 19th European Conference on Pattern Languages of Programs*, 22.

Riedel, J. C., & Hauge, J. B. (2011, June). State of the art of serious games for business and industry. In *2011 17th International Conference on Concurrent Enterprising*, 1-8.

Rivera-Vazquez, J. C., Ortiz-Fournier, L. V., & Rogelio Flores, F. (2009). Overcoming cultural barriers for innovation and knowledge sharing. *Journal of Knowledge Management*, *13*(5), 257-270.

Rouiller, J. Z. and Goldstein, I. L. (1993), 'The relationship between organisational transfer climate and positive transfer of training', *Human Resource Development Quarterly*, *4*(4), 377–90.

Ruohomäki, V. (2003). Simulation gaming for organizational development. *Simulation & Gaming*, *34*(4), 531-549.

Santos, J. R. A. (1999). Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of extension, 37*(2), 1-5.

Savi, R., von Wangenheim, C. G., & Borgatto, A. F. (2011, September). A model for the evaluation of educational games for teaching software engineering. In *2011 25th Brazilian Symposium on Software Engineering*, 194-203.

Schmidheiny, S., & Timberlake, L. (1992). *Changing course: A global business perspective on development and the environment*, *1*

Schultz, C. P., & Bryant, R. D. (2016). *Game testing: All in one*. Stylus Publishing, LLC.

Schumann, P. L., Anderson, P. H., Scott, T. W., & Lawton, L. (2014). A framework or evaluating simulations as educational tools. In *Developments in Business Simulation and Experiential Learning: Proceedings of the Annual ABSEL conference*, *28*

Simpson, E. J. (1966). The Classification of Educational Objectives, Psychomotor Domain.

Slavin. R. E. (1997). Research on cooperative learning and achievement: A quarter century of research. Paper presented at the Annual meeting of Pedagogical Psychology. September, Frankfurt.

Smith, E. A. (2001). The role of tacit and explicit knowledge in the workplace. *Journal of knowledge Management*, *5*(4), 311-321.

Steiner, C., Hollins, P., Kluijfhout, E., Dascalu, M., Albert, D., Westera, W., & Nussbaumer, A. (2015). Evaluation of serious games: a holistic approach. In *ICERI2015: 8th annual International Conference of Education, Research and Innovation*, 4334 4342.

Susi, T., Johannesson, M., & Backlund, P. (2007). Serious games: An overview.

Sutcliffe, M. (2002). Simulations, games and role-play. *The handbook for economics lecturers*, 1-26.

Tamkin, P., Yarnall, J., & Kerrin, M. (2002). *Kirkpatrick and Beyond: A review of models of training evaluation*. Brighton, England: Institute for Employment Studies.

Van den Bossche, P., Gijselaers, W. H., Segers, M., & Kirschner, P. A. (2006). Social and Cognitive Factors Driving Teamwork in Collaborative Learning Environments. *Small Group Research, 37*(5), 490-521. doi:10.1177/1046496406292938

Van den Bossche, P., Gijselaers, W., Segers, M., Woltjer, G., & Kirschner, P. (2011). Team learning: building shared mental models. Instructional Science, 39(3), 283-301.

Van Pelt, S. C., Haasnoot, M., Arts, B., Ludwig, F., Swart, R., & Biesbroek, R. (2015). Communicating climate (change) uncertainties: Simulation games as boundary objects. *Environmental science & policy*, *45*, 41-52.

Veldhuis-Diermanse, A. E. (2002). *CSC Learning?: participation, learning, activities and knowledge construction in computer-supported collaborative learning in higher education*.

Vermunt, J. D. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher education*, *31*(1), 25-50.

Vermunt, J.D. (1992). Leerstijlen en sturen van leerprocessen in het hoger onderwijs. Naar procesgerichte instructie in zelfstandig denken.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes* Harvard university press.

Webb, N.M., & Palincsar, A. S. (1996) Group processes in the classroom. *Handbook of educational psychology*, 841-873.

Whalen, K. A., Berlin, C., Ekberg, J., Barletta, I., & Hammersberg, P. (2018). 'All they do is win': Lessons learned from use of a serious game for Circular Economy education. *Resources, Conservation and Recycling*, *135*, 335-345.

White, D. D., Wutich, A., Larson, K. L., Gober, P., Lant, T., & Senneville, C. (2010). Credibility, salience, and legitimacy of boundary objects: water managers' assessment of a simulation model in an immersive decision theater. *Science and Public Policy*, *37*(3), 219-232.

Wildman, S., & Reeves, M. (1996). The utilization and evaluation of a simulation game in pre-registration nurse education. *Nurse education today*, *16*(5), 334-339.

Worldometers.info. (2019). *World Population Clock: 7.7 Billion People (2019) Worldometers*. [online] Available at: http://www.worldometers.info/world population/ [Accessed 23 Jan. 2019].

Appendix

Appendix I. Questionnaire

Questionnaire for the evaluation of the <name Serious Board Game>

Please, help us improve the game answering the following questions. All information is collected anonymously and will be used only in a summarized way in the context of this game evaluation. To fill in the questionnaire takes approximately 5 minutes.

Demographic information	
Place and date	
1. Which of the following groups do you associate with your work?	☐ business and industry ☐ civil society ☐ NGOs ☐ policy ☐ science ☐ student ☐ other, please specify: _____
2. Work experience in this field:	☐ 0 – 1 year ☐ 2 – 3 years ☐ 4 – 6 years ☐ 7 – 10 years ☐ > 10 years
3. During the game I had the role of:	☐ Government ☐ Industry ☐ Market / Bank
4. Are you working in the food sector?	☐ Yes ☐ No (please, skip question 5)
5. Do you think there is a link between the food sector and the subject of the game?	☐ Yes, because_____ ☐ No

6. Did you know the other participants?	☐ Yes, I knew all the participants ☐ I knew some of the participants ☐ No, I did not know any of the other participants
7. In general, what grade would you give the game? (1 = very poor, 10 = excellent)	1 2 3 4 5 6 7 8 9 10 ○ ○ ○ ○ ○ ○ ○ ○ ○ ○

Please, **select an option** according to how much you agree or disagree with each statement below.

Player Experience						
No.	Statement	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
8	I consider myself as an expert on the subject of the game					
9	The game provides new challenges (offers new obstacles, situations or variations) at an appropriate pace					
10	Opinions and ideas of participants are verified by asking each other critical questions					
11	This game is appropriately challenging for me					
12	Information from participants is complemented with information from others					
13	I had the impression that other participants shared their knowledge					

14	The participants tend to handle differences of opinions by addressing them directly					
15	Participants elaborate on each other's information and ideas					
16	I would play a serious game again					
17	I had the feeling that nobody was listening to me					
18	During the game I shared all relevant information and ideas I had					
19	I had the impression that other participants had less expertise on the subject than I have					
20	The game does not become monotonous as it progresses (repetitive or boring tasks)					
21	I felt good interacting with other players during the game					
22	In the game I had the opportunity to share my vision					
23	I would recommend this game to my colleagues/friends					
24	Dialogues with other participants lead to new insights to me					
25	Comments on ideas are acted upon					
26	I was able to interact with other players during the game					
27	Participants were listening carefully to each other					

Please, **select an option** according to how much you agree or disagree with each statement below.

Learning						
No.	Statement	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
28	The game contributed to <*verb related to the level of the learning goal*> <goal/concept>.					
29	The game contributed to *understand* the need of circular resource recovery					
30	The game contributed to					
31	The game contributed to					
32	The game contributed to					

Transfer						
No.	Statement	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
33	After playing the game, I am motivated to start a discussion/dialogue with other participants					
34	I will definitely spend time thinking about how to use the knowledge that I have gained					

For which audience would you recommend this game?

Do you have any recommendations where we can play this game?

_____

_____

My general suggestions for improving this game are:

1._____

2._____

3._____

The strong aspects of this game are:

1._____

2._____

3._____

Do you have any further comment?

_____

_____

Thanks a lot for your contribution!

Appendix II. Questionnaire with additions

Questionnaire for the evaluation of the Serious Board Game

Please, help us improve the game answering the following questions. All information is collected anonymously and will be used only in a summarized way in the context of this game evaluation. To fill in the questionnaire takes approximately 5 minutes.

Demographic information	
Place and date	
1. Which of the following groups do you associate with your work?	☐ business and industry ☐ civil society ☐ NGOs ☐ policy ☐ science ☐ student ☐ other, please specify: _____
2. Work experience in this field:	☐ 0 – 1 year ☐ 2 – 3 years ☐ 4 – 6 years ☐ 7 – 10 years ☐ > 10 years
3. During the game I had the role of:	☐ Government ☐ Industry ☐ Technology provider
6. Did you know the other participants?	☐ Yes, I knew all the participants ☐ I knew some of the participants ☐ No, I did not know any of the other participants
7. In general, what grade would you give the game? (1 = very poor, 10 = excellent)	1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 ○

8 How often did you play serious games?	☐ Never ☐ Rarely, once or twice ☐ Quite often, more than two times

Please, **select an option** according to how much you agree or disagree with each statement below.

Player experience						
No.	Statement	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
9	The game provides new challenges (offers new obstacles, situations or variations) at an appropriate pace					
10	Opinions and ideas of participants are verified by asking each other critical questions					
11	This game is appropriately challenging for me					
12	Information from participants is complemented with information from others					
13	The participants tend to handle differences of opinions by addressing them directly					
14	Participants elaborate on each other's information and ideas					
15	I would play a serious game again					
16	During the game I shared all relevant information and ideas I had					
17	The game was difficult to understand and play					

18	I felt good interacting with other players during the game					
19	I would recommend this game to my colleagues/friends					
20	Comments on ideas are acted upon					
21	I was able to interact with other players during the game					
22	Participants were listening carefully to each other					
23	I think the explanation of the game missed some essential information					

Please, **select an option** according to how much you agree or disagree with each statement below.

Learning						
No.	Statement	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
24	The game contributed to					
25	The game contributed to					
26	The game contributed to					
27	The game contributed to					
28	The game contributed to					
29	The game contributed to					

Transfer						
No.	Statement	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
30	After playing the game, I am motivated to start a discussion/dialogue with other participants					
31	I will definitely spend time thinking about how to use the knowledge that I have gained					

For which audience would you recommend this game?

_____

Do you have any recommendations where we can play this game?

_____

My general suggestions for improving this game are (think also about the explanation of the facilitator):

1._____

2._____

3._____

The strong aspects of this game are:

1._____

2._____

3._____

Do you have any further comment?

_____

_____

Thanks a lot for your contribution!