Gaming to Make Sense of Risks: Toward a Thicker Description of Game-Based Training

Casper Harteveld

Abstract: In the past decades we have seen an increased interest in the use of games beyond entertainment. Although a few successes are known and valuable insights are provided on the use of games for impact, the field is in dire need of comprehensive and rigorous studies. To fill this gap in the area of game-based training, in this chapter the study and results of a unique game called *Levee Patroller* are discussed. In this game, players will need to make sense of levee risks. The study's objectives were to design and implement an innovative game-based training intervention that would allow to gather empirical evidence as well as a substantiated understanding of how participants experienced the training and played the game. A mixed methods research design was employed using questionnaires, tests, and game data, among others. With the results a "thicker description" was retrieved from which it became clear that games can be a potentially powerful tool to enable players to make sense of phenomena, but that the full potential has not been reached. Future research should continue to employ mixed methods research designs to capture the complexity inherent in playing games and to bring game research to another next level.

1. Introduction

Along with the rise of digital games over the past decades came an increased interest for using games beyond entertainment (Harteveld, 2011). So it happened that games have appeared to make children knowledgeable about diseases as well as games that help to make computers smarter. Although a few successes are known, much research–in particular about educational games–seems to suggest little evidence for games' advantages (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Hays, 2005; O'Neil, Wainess, & Baker, 2005; Sitzmann, 2011; Young et al., 2012). Such evidence makes clear that we need to speak of "the rise of a *potential* powerful tool." Gaming has potential, theoretically and based on some of the "hints" from literature, but we need to figure out how to utilize and proof that potential.

The existing literature suggests that more studies are needed that investigate the *effective* design and use of games. Although extant research has taught us a number of lessons and the field seems to have learned from the mistakes of the past, we still have no clear idea of how games are produced that provide reliably pre-specified objectives (Tobias & Fletcher, 2012).

In addition, the field is especially in dire need of comprehensive and *rigorous* studies, that is, studies that go beyond anecdotal, descriptive, or judgmental evidence that does not suffer from methodological flaws. Such studies require *innovation* too, because we may not be able to make use of games effectively without it and/or capture what impact they have.

To contribute to this emerging field, the case of *Levee Patroller* was investigated (Harteveld, 2011; Harteveld, 2012). This unique game was developed in 2006 and its name refers to the game's target group. Levee patrollers are considered the "eyes and ears" of the Dutch water authorities, which are organizations that are responsible for the water quality, quantity, and safety in the Netherlands. They inspect levees, the artificial and natural barriers that protect a region from flooding, and report any risks they encounter. Much similar to the actual practice, players have to find all virtual failures in a region and report them. If they do not find the failures in time or report them incorrectly, it could result in a levee breach that floods the whole virtual region.

This case was investigated for two reasons. First, it may be a unique game, but it is not alone. Many similar game-like digital technologies have been developed in the past decade, such as for training first responder response to hazardous materials and triaging patients

during a crisis. These technologies all share in common that they are situated in the same domain, one of *safety and crisis response*. They further attempt to bring forth the same value and by similar means. They aim for *knowledge* about risks and achieve this by means of *sensemaking*, which is roughly defined as a process by which people give meaning to phenomena. Finally, they even use a similar type of game genre. Each can be considered a 3D simulation. Therefore, investigating *Levee Patroller* helps to shed light on a particular specialization within the emergent use of games for serious purposes: the use of games to make sense of risks.

Second, *Levee Patroller* might be unique; it provides for a unique opportunity to contribute to advancing the field. Little is known about the use of games in the domain of safety and crisis response, a domain for which gaming has an incredible potential. Unlike other closely related technologies, this game has been fully developed to facilitate many hours of training and it found an actual application, as five water authorities participated in its development and wanted to build a curriculum around it.

The objectives behind the investigation were two-fold. The first objective relates to the dire need for evidence about the effectiveness of games. This objective was to design and implement an innovative game-based training intervention and evaluate its effectiveness in a comprehensive and rigorous manner. The following questions are associated with this objective:

1. What is the effectiveness of the training with *Levee Patroller*?
2. What factors contribute to its effectiveness?

Because very little is known about game-based training and in particular regarding the domain of safety and crisis response, the second objective was to develop a substantiated understanding of what makes a game successful in training practitioners to make sense of risks. By considering the following questions such understanding would be developed:

3. How do participants experience the game-based training?
4. How do participants play the game?

However, what was really aimed for in this investigation was to establish a *thick description* (Geertz, 1973) of a game-based training. The study did not only aim for measuring the results, but also for providing a context from how these results were established. Because a mix of methods and methodologies were used to get this description, we could speak of establishing a “thicker description.”

In this chapter, this thicker description of *Levee Patroller* will be revealed piece by piece. First, more background regarding the development of the game and the design of the game is provided. What follows is a description of the comprehensive approach in answering the afore-mentioned research questions. Before the results are revealed, a detailed description is provided on how the game-based training was implemented and experienced. This will give the appropriate context and the necessary nuances for interpreting the results. This chapter ends by concluding that games are indeed a potential powerful tool, and that we now have a much desired rigorously studied example of a game that illustrates this. The study and its results can be used to take game research to another next level.

2. Background

The game *Levee Patroller* was developed in the Netherlands, a country where levee inspection has deep historical roots. For a long time the water authorities were the only accepted authorities in the Netherlands, and this was out of necessity. Only by collaborating could inhabitants face the floods risks, which are considered a risk greater than all other risks combined in this country. For example, between 1750 and 1800 alone, there were 152 floods (Rijkswaterstaat, 2011). The oldest water authorities were established in the 13th century.

Their influence extended much beyond dealing with flood risks. They had a social function and judicial power.

Over time, the number of water authorities has been drastically reduced as well as their responsibilities. In 1850, the number of known authorities surmounted to 3500 and nowadays there are 27. Authorities received one specific dedicated task: to manage the water system in their region. This management involves three aspects: water quantity, water quality, and water defenses. Levee inspection pertains to the latter. This inspection happens regularly and during emergencies to determine the state of the levees.

Although levee inspection has been around since the Middle Ages, it was not until 2001 when the ad hoc nature of how the water authorities dealt with emergency situations was replaced by an organized structure within these institutions that carefully prepares, plans, and coordinates the prevention and after care of calamities using standard procedures and information technology. This was done as a response to the floods in 1993 and 1995, from which it was concluded that a better organization was needed, with practitioners who are prepared to deal with the unexpected. The *Levee Patroller* game was developed within the context of the desire to professionalize how the Dutch water authorities managed flood risks.

Another reason for its development is that the patrollers have to deal with rare but disastrous failures. Levee failures hardly occur and it is difficult to get practical experience. In fact, despite its “virtuality” the game provides the only means to get experience in finding and reporting levee failures. The consequences of a potential failure are furthermore too large, and technology can only assist with inspecting levees to some extent. For example, there are simply too many levees to equip them all with sensors. Therefore, personnel is required that knows what the risks are and how to deal with them.

Although levee inspection is unique to the Netherlands, similar inspection practices, with water authorities and patrollers, exists elsewhere too. Also, the Netherlands is by far not the only country with flooding issues, as the disaster of Hurricane Katrina illustrated, and with climate change and advanced urbanization more flooding issues can be expected across the world in the nearby future. The game is for this reason potentially applicable beyond the Netherlands, and it has in fact been played by practictioners all over the world. However, the development was primarily initiated due to the needs of the Dutch water authorities and they were the clients for which the game was originally developed.

2.1 Learning Objectives

The exact roles and responsibilities differ per water authority, and also how they are implemented. The responsibilities of patrollers are generally to: 1) find failures; 2) report signals that make up a failure; 3) communicate reports to others; 4) diagnose the situation; and 5) take measures when necessary. Patrollers report their findings to a coordinating field office called *Action Center*. They collect all the information and provide further instructions to the people in the field. In developing the game, five learning objectives were identified based on these responsibilities: observing, reporting, assessing, diagnosing, and taking measures.

These objectives indicate that the game involves knowledge: knowledge about recognizing failures and how to deal with them. In the end, the game is especially about teaching an ability to make sense of (virtual) risks, which involves technical skills called *sensemaking skills*. Sensemaking is a process that leads to the (re-)construction of knowledge (Dervin, Foreman-Wernet, & Lauterbach, 2003; Weick, 1995). This process inlcudes information gathering, processing, and then taking action, and happens especially when people are challenged and want to create order out of chaos.

The game achieves its objectives by engaging players in a process with many challenges that they have to make sense out of and by guiding the player’s meaning construction of these challenges. The game provides a structure of what failures exist and how they need to be recognized and dealt with. Despite this steering (or sensegiving) by the game, the player’s

meaning construction may still widely differ, because sensemaking processes do not occur in a vacuum. A player's history, culture, identity, and other factors play a role in how knowledge is (re-)constructed.

The specific kind of knowledge the game is helping to establish are:

1. *Inspection concepts and vocabulary*: Players learn what concepts are relevant to levee inspection and how to label them, from failure signals to failure mechanisms;
2. *Mental model formation of failures*: Players learn what failures exist, where they can occur, how to recognize them, and what failure mechanisms and measures relate to them; and
3. *Mental simulation of failures*: Players learn how a failure develops over time. Based on this they can develop expectations.
4. *Inspection protocol*: Players learn a set of rules and procedures that must be followed during an inspection process.

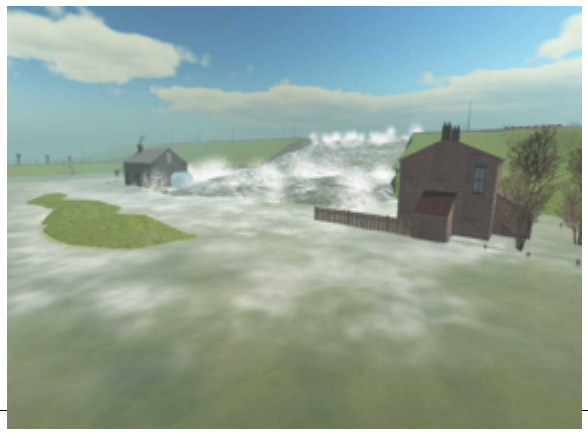
The game does not only help to establish this knowledge, it also teaches them how to apply this in the situated context of dealing with levee failures. This is why in the end the game is about teaching sensemaking skills. It is really about teaching players the ability to apply the desired for levee inspection knowledge.

Although it was not primarily developed with this purpose in mind, the game may have an impact on *communication* too. It arguably provides for a *common vocabulary* and *shared experience* that will make communication easier between the different actors involved in the inspection process. Speaking the same language and having a similar experience to draw upon will more likely lead to sharing the same meaning. For these reasons, despite that the game was developed as a single player game, it can still foster improved collaborative decision-making. This is important to stress because levee inspection–and many other organizational processes–do not happen in a vacuum. Multiple actors are involved and a game should acknowledge this multi-actor complexity of reality.

In short, these concepts relate to each other as follows: the game enables sensemaking; this sensemaking leads to knowledge; knowledge enables the development of (technical) skills; and having learners acquire knowledge and skills from a similar sensemaking process may have an impact on communication. This has been the educational philosophy behind the development of *Levee Patroller* through which an attempt was made to make an impact in the field of levee inspection.

2.2 Game Design

Using information from the real world and keeping the educational purpose in mind, an interdisciplinary team started to develop a game that is now referred to as *Levee Patroller*. The game was developed in close collaboration with five water authorities and using the Unreal Engine 2.0, a game engine that makes it easier for developers to create games. The first version was developed in nine months.



(a) Levee failure (b) Flooding

Figure 1. Two screen shots of the game *Levee Patroller*

In this single player game, players can walk freely around a 3D environment–the purpose of the movement being to find all potential failures and appropriately deal with them using the tools of the game (e.g., handbook, map, etc.). Dealing with failures requires players to *recognize* them. Subsequently, players have to *report* the location of the problem and the signals that constitute the potential failures and their characteristics. Next the player must communicate the findings to a field office and *assess* the severity of the situation. Failures can change over time so players need to return to the problem locations to determine if the failure potential worsened. If it did, players make another report to a field office. If the failure becomes critical, i.e., it is likely a levee breach will occur, players have to indicate this to the field office. After *diagnosing* the situation, which involves determining what mechanism is causing the impending failure, players can *take action* and implement a stabilizing measure. If they do not implement this in time or use the wrong technique, a catastrophic failure and flooding results. Throughout and at the end of an exercise, players receive feedback on how they recognized, reported, assessed, diagnosed, and took action against failures.

The game has different regions with each specific characteristics and includes a variety of failure types. Figure 2 illustrates one of the failures and the flooding that could be the result of it when it is not appropriately dealt with. A training level was created to teach players how to navigate the 3D environment and teach them the various tools they can use to deal with failures.

Table 1. Overview of *Levee Patroller*'s scoring system

Criterion	**Description**	**Learning Objective**
Observed failures	The player indicates that he or she has found a failure	Observing
Location accuracy	The player specifies the location of a failure	Reporting
Observed signals	The player reports what signals are part of a failure	Observing
Reporting accuracy	The player fills out a report for each signal	Reporting
Assessment accuracy	The player makes an estimate of how severe the situation is	Assessing
Diagnose accuracy	The player determines the failure mechanism behind a failure	Diagnosing
Measure effectiveness	The player takes an action to prevent a failure from becoming worse	Taking measures

An important design decision was to use the learning objectives to design the activities in the game and as criteria for a game score. Table 1 provides an overview of the criteria in the game, what they mean, and what learning objective is associated with each. There are two criteria for observing, one for observing the failure itself and one for recognizing the signals. A failure could consist of one or more signals, so it is important to make this distinction. For reporting we included two criteria as well after consultation with the water authorities. They mentioned that reporting the location is an important and often overlooked aspect in the inspection process and they wanted to stress this importance in the game. By weighting the

criteria and summing the weighted scores for each failure in a scenario, a *total score* is calculated. This total score is presented as a percentage of the maximum score that a player can achieve in a scenario.

Immediately after its release one of the water authorities enthusiastically started to make use of it, on their own initiative. They created a special room in their office building and called this the “game room”. However, the facilitators were not knowledgeable about the game and the time spent on playing was far too little to get much out of it. Another use was as part of the traditional levee inspection instruction courses. Even less time was dedicated to play the game here. The game was basically used for visualization there. In conversations with the authorities it became clear that they had difficulty in thinking about how to utilize the game effectively. Therefore, part of the research concerned designing a training that would make appropriate use of the game.

3. Training/Evaluation

To study *Levee Patroller* in a comprehensive and rigorous manner a *training/evaluation* was set up. It is deliberately called a training/evaluation because as part of this study it was necessary to think of what would be considered an effective training as well as how this effectiveness could be measured. For designing and implementing the training/evaluation, the following ten principles were kept in mindd based on a review of the literature:

1. *Rome was not built in a day*: Although gaming has a long and rich history, game-based training has not been comprehensively studies until recently. We cannot expect that game-based training works most optimally right away. An understanding of what works and what does not is more fruitful than the evidence itself;
2. *No comparison of apples and oranges*: Although comparisons can be useful, they are not always possilbe or sensible. In addition, game-based training deserves further investigation in its own right;
3. *More than the tip of the iceberg*: A comprehensive examination needs to go far beyond the standard questionnaires usually associated with game-based evaluations. Proof that it works requires a rigorous evaluation.
4. *The proof of the pudding is in the eating*: Players learn from playing the game, but often the gameplay is treated as a black-box. By not considering the gameplay much valuable information is lost.
5. *The icing and the cake*: In evaluating a game not only obvious design elements should be considered, such as its graphics and controls. The complete design needs to be taken into account. This requires a full understanding of the design by evaluators.
6. *Ain’t nothing like the real thing*: If a game is designed for a specific target group, then it should be played with that target group. Playing the game with, for example, students may lead to different outcomes.
7. *Practice makes perfect*: Training requires practice–also with digital games. This means a game needs to be played more than once to see its effects and facilitators/evaluators need to consider this in their setup.
8. *Big fish in a small pond*: The impact of a game involves much more than the quality of the game, although that is a major part (hence it is the big fish). Facilitation, documentation, and other contextual variables matter too.
9. *See the big picture*: Evaluation needs to be comprehensive, but it impossible to take every aspect into account and to consider those that are in a detailed manner. It is a choice of breadth over depth. This also means the research should not focus all its attention on one or two aspects.
10. *It takes two to tango*: It takes two to tange, but it takes three to design a meaningful game. It also takes three to evaluate a game. Game evaluation needs to occur in an interdisciplinary fashion by taking multiple perspectives into account.

These evaluation principles declared the focus, scope, and assumptions behind the design of the evaluation and the training. It is important to note that the training and evaluation goals are slightly different (i.e., with the training the trainees were the major concern and with the evaluation the research results). This led to some tensions in the study's design and execution. Tensions were resolved by minimizing the negative implications of each design choice.

3.1 Evaluation design

In a nutshell, the main hypothesis behind the game-based training with *Levee Patroller* is that after the training participants will have increased their levee inspection knowledge and skills and are able to communicate better with each other. For achieving a thicker description of game-based training, the first objective was to find evidence for this hypothesis. The second objective was to retrieve a substantiated understanding. Proofing that it works is just as useful as understanding why it works. To achieve both objectives, the strategy was to ground the evaluation in *mixed methods research* that combines a quantitative approach to retrieve the evidence and a qualitative approach to retrieve the substantiated understanding. In the evaluation, the mixed methods approach is embedded within that of a *quasi-experimental design* (Creswell & Clark, 2007), and more specifically a *one-group pre-test-post-test design* (Cook & Campbell, 1979). This design is consistent with the afore-mentioned principles–by focusing on the game and doing this in a comprehensive manner (Principles 1-3).

Its unit of analysis concerned the (individual) players with as target group the actual levee patrollers (Principle 6). The main outcomes to evaluate the game's impact were *judgments*, *knowledge perception*, and *sensemaking performance*. Judgments are about how players appreciated the game/training; knowledge perception refers to players' self-assessment on the various learning objectives; and sensemaking performance is about how practitioners are able to deal with failures. *Communication* and affective learning outcomes, such as perceptions regarding levee inspection and a possible heightened awareness, were considered secondary outcomes. As the game-based training was an individual training, the proxy of vocabulary usage was considered to measure communication, and assume that with increased similarity, better communication would be the result. The following working hypotheses were used to guide the evaluation for measuring the main and secondary learning outcomes:

Hypothesis 1: *Post-training knowledge perception and sensemaking performance will be higher compared to the pre-training knowledge perception and sensemaking performance, respectively.* This is the main hypothesis referered to earlier. It includes a self-assessment component and a test-assessment component.

Hypothesis 2: *Knowledge perception does not correlate strongly with sensemaking performance.* Existing evidence suggests that self-assessment and learning are only moderately correlated (Sitzmann, Ely, Brown, & Bauer, 2010). If this were to be true for *Levee Patroller* as well, it would be another indication why it is important to go beyond the tip of the iceberg (Principle 3).

Hypothesis 3: *Post-training sensemaking performance on virtual pictures is higher compared to real pictures.* For measuring sensemaking performance a test was developed and in this test participants will have to make sense of real and virtual pictures. This allows for measuring a *transfer* effect. However, the virtual pictures came straight from the game and the real ones the participants may have never seen. Therefore, a difference is to be expected between the two types of pictures.

Hypothesis 4: *Game judgments correlate with knowledge perception and sensemaking performance; participants who evaluate the game higher will have a higher knowledge perception and sensemaking performance.* If players have fun, are engaged with the game, and conceive that they are learning from it, it is likely that they are the ones who will benefit

from the game (either perceived or through more objective measures). This has to do with a feedback loop that motivates players to learn more (Garris, Ahlers, & Driskell, 2002).

Hypothesis 5: *Post-training vocabulary use will resemble the game's vocabulary closer compared to pre-training vocabulary use.* By playing the game participants will acquire or adapt the game's vocabulary. If this were to be true, this is evidence for a common vocabulary, which will arguably lead to an improved communication.

Various other aspects were measured based on a full understanding of the game's design (Principle 5); taking multiple perspective into considersation, such as gameplay, learning, and the domain of levee inspection (Principle 10); and to ensure for a comprehensive evaluation (Principles 5, 9, and 10). It was hypothesized as follows on how some of these additional measures would have an impact on the main outcomes:

Hypothesis 6: *The number of exercises played moderates the results on the main outcomes; participants who play more will have higher results one the main outcomes.* This hypothesis relates to the principle "practice makes perfect" (Principle 7).

Hypothesis 7: *Game scores moderate the results on the main outcomes; participants with higher game scores will have have higher results on the main outcomes.* In *Levee Patroller* the learning objectives are directly coupled to the game scores. This means it is arguable that if someone is better in playing the game, that person has more knowledge and skills about levee inspection.

Hypothesis 8: *Computer skills and game skills moderate the game scores; participants with higher computer and game skills will have higher game scores, respectively.* Players will first have to learn how to play the game before they can learn from it. Considering that *Levee Patroller* is a digital game, players will need to have considerable computer skills. Participants who have the necessary skills can pick up the game faster and are better able to focus on the content. For participants with significant experience in playing games this will arguably have an even stronger effect.

Hypothesis 9: *Game attitude moderates the game judgments; participants with a higher game attitude will have a higher game judgment.* Certain participants may like to play games and others may not. It can be expected that a participant's predisposition has an impact on how they experience and value the game.

Hypothesis 10: *Younger participants achieve higher game scores compared to older participants.* A frequent re-appearing argument is the division between people who have grown up playing digital games and those who have not (Prensky, 2001). People growing up with games have different expectations and skills, and it can be expected that they are better in playing games. This, in turn, may impact what they gain from playing (Hypothesis 7).

The following mix of qualitative and quantitative methods were used to measure and validate the outcomes and the likely moderating variables:

Pre- and post-questionnaire: Before and after the training, participants made a self-assessment of their knowledge and attitudes toward levee inspection. The pre-questionnaire was further used to gather contextual variables, such as age and game attitude, and the post-questionnaire to determine how participants judged the training.

Pre- and post-sensemaking test: To determine the sensemaking performance, participants needed to make sense of pictures with virtual and real failures before and after the training. Open questions were used to retrieve the purest answers, which were not influenced by the researcher's wording or categorization, and to see an impact on communication by considering vocabulary usage.

Game questionnaire: Participants had to answer a small questionnaire based on a number of closed and open questions after every exercise. This was used to understand how participants experienced a particular exercise and see how their experience with the game

might change over time. This method was included to understand what is happening while participants are playing the game (Principle 4).

Game data: Each exercise resulted in game data. This game data consists of quantitative and qualitative data of how the participant played an exercise. With this data a reconstruction was made of how participants made sense of virtual failures. This method was also included with the aim of understanding what players do (Principle 4).

A number of additional measures were used to triangulate the results from these data, such as interviews and discussions (Harteveld, 2012). In general, these additional measures validated the results from the main methods.

3.2 Training design

In designing the training, the evaluation design was kept in mind, as well as any practical considerations resulting from how the training could be successfully implemented with the target group. Further requirements were to make sure participants would get sufficient opportunity to practice with the game (Principle 7) and that the game would be supplemented with additional instructional methods (Principle 8). The largest concern was whether participants would play at all. To encourage players to play, notions from *cognitive load theory* (Plass, Moreno, & Brünken, 2010) and some common sense ideas about the willingness and commitment of participants were considered.

Of importance was to make sure that the “extraneous cognitive load” of learning how to play the game would be reduced and think of ways of how participants could better learn how to learn (i.e., “germane cognitive load”). *Cognitive overload* (Mayer & Moreno, 2003) was to be avoided, as this meant the participants would not be able to process the information from the game and therefore not learn from it. To achieve this, a meeting to play the game under guidance was deemed necessary. It also seemed better to spread out the practice to facilitate processing of the information in a piecemeal manner and provide for opportunities to reflect on the material in between and over a longer period of time. Additional instructional support, such as a website and a manual, would help participants in learning the material and how to play the game.

In order for the game to make an impact, the participants would have to play more than once. Prior to the training it was a concern if participants were willing to spend their free time on the training. Also, many amongst the target group were volunteers and their availability would be limited during the day. However, employees would arguably be less interested in anything offered after work hours. Flexibility seemed key here and for this reason it was decided to have the participants *play at home*. The optimal length was set to three weeks; anything more seemed psychologically too long to commit to.

Having people play at home has its disadvantages, and the many online learning programs confirm that the majority of the learners do not complete them (Meister, 2002). The following was done to increase committment of the participants. First, participants would physically attend a *start-* and *end-meeting*. The questionnaires and tests were filled out at these meetings, ensuring that everyone completed them. The end-meeting further ensured for a clear deadline to finish all the assigned exercises, whereas the start-meeting made sure that the participants went home knowing what to do. In between the two meetings, they would play at home. The meetings seemed the perfect compromise in giving the participants the flexibility to play at home.

Second, *weekly assignments* were included. Every week participants received a code to unlock the next two exercises of the game. The assignments enforced to play the game in a structured manner and spread out over the training, as suggested by cognitive load theory. It also provided an inobtrusive reminder to participants to play the game and for the facilitator to stay in touch with the participants.

A special research version of *Levee Patroller* was developed with eight exercises. The first and last exercise were scheduled to be played at the start- and end-meeting, respectively. The other six exercises were assigned to be played at home, two per week. The exercises increased in difficulty by adding more (severe) failures and ends with a spectacular grand finale. For example, the first exercise includes two failures, none of which are critical. The last exercise includes five failures and four of those are critical. Other considerations in designing this training program were internal validity and variety. It was made sure that players would play the exact same exercises and in the same order. Participants were unable to replay previous exercises. The variety consideration was made to keep the training interesting and engaging. This was accomplished by varying the types of failures, the environments, and the weather.

4. Setup and Implementation

Three water authorities agreed to implement the training/evaluation. The setup differed per authority, in terms of training administration, recruitment, location, support, compensation, and its premise. All of these items were decided upon in negotation with the participating water authorities. These differences in setup and the authorities itself seemed to be influential in how the training was experienced and are for this reason detailed here. The authorities are heretofore called Organization A, B, and C.

The first water authority, Organization A, saw the training as an opportunity to revamp its relationship with its patrollers. Although they are the ones who enthusiatistally started to use the game and created a game room, and have been propelling expansions of the game, they had not organized many activities for years. Out of 178 levee patrollers associated with Organization A, 37 participated. Because this authority serves one of the most densely populated and industrialized regions, it can be characterized as “innovative, industrial, and inactive.”

Organization B was convinced about the game’s usefulness but did not know how to implement a game-based training. For them the training was an opportunity to find out if this was even possible. Their difficulty in finding a way to implement a training was due to the size of their organization. This organization has 650 patrollers spread out over five areas within the region the authority serves. For this pilot they did not want to reach out to everyone. Restricited by the room size of the training location, we sampled about 15 participants from each area and recruited in total 77 participants. In stark contrast with Organization A, this water authority was very active with their patrollers and had a good vision and strategy for the future on how to professionalize their levee inspection. It can be characterized as “ambitious, big, and clueless.”

Figure 2. One of the meeting sessions

The third authority, Organization C, still had to be convinced about the usefulness of the game and thought along on how evidence could be retrieved about its use. They further invested much in developing inspection protocols and procedures in order to professionalize their levee inspection. They wanted to convey the seriousness of this training and insisted that this training was made compulsory to one of their areas. This premise made a huge difference. Participants–in particular the volunteers–disliked the fact that it was made compulsory and the training started off very negatively, with a few participants unwilling to participate. Including the unwilling participants, in total 33 participants out of the 300 patrollers associated with this organization participated. This authority was located in a rural area in the Netherlands and most of its participants lived close to levees. Overall, the participants seemed more knowledgeable about levees compared to the participants of the other water authorities. Organization C can be characterized as "structural, critical, and rural."

The total number of participants came down to 147. These participants were relatively old (M = 47.6; SD = 12.1); were practically all male; had a mixed education and diverse occupations; had little failure and game experience; and had little computer skills. Some did not even own a computer. These participants received a laptop for the duration of the training.

Of this number, 5% dropped out of the training and for various reasons–predominantly a dislike regarding the game. However, in general the game-based training can be conceived as successful. A large majority (80%) played at least five out of six exercises at home, which is a participation rate that exceeded expectations. On average it took players an hour to play each exercise. Adding all the other training activities, participants spent about two full work days on the training.

Although successful, many problems occurred throughout the training that need to be noted. For various reasons, about 11% of the data went missing, despite having built in redudancy for saving the files locally on a server. The initial training setup at Organization A needed significant improvements. Amongst others, the time of the start-meeting was too short and a number of small game elements led to much frustration. These improvements were made with the goal of training in mind.

5. Results

In addition to the high participation rate, there are other findings for concluding that the game can be considered effective. In terms of the analysis, the items on the questionnaires were inspected individually first. Composite items were created using a principal components analysis after that. From this the components pre-knowledge perception (α = .928), game attitude (α = .662), post-knowledge perception (α = .909), and judgment (α = .914) have been derived.

The pre- and post-sensemaking tests were analyzed using content analysis. With this content analysis, the types of answers and the accuracy was checked. For the types of answers, the literal answers were used. Only in case of spelling errors corrections were made. For example, if a participant answered "stones" and another "rocks," these were considered two separate answers. However, "stone" and "stones" were considered similar. In this manner, vocabulary usage was determined.

For accuracy the logic and vocabulary of the game was used as a normative model. A score was considered *very accurate* if it was literally similar to the content of the game (e.g., "pitching stone"); *accurate* if it was closely similar to the content of the game, a synonym, or a proper alternative (e.g., "pitching rock"); *slightly accurate* if it was not necessarily wrong but descriptive or when vague language is used (e.g., "stones"); and *inaccurate* if the response was simply incorrect or too vague. Test scores were calculated by summing the scores across all answers–in total and for the virtual and real pictures separately.

For the open questions on the game questionnaires pattern coding was employed and through this the major patterns were revealed. For the closed questions with Likert items, it was observed how player ratings changed over time. The raw game data was inspected per participant over time and in total.

5.1 Empirical evidence

It turns out that the game-based training impacts strongly and positively participants' knowledge perception, $t(111) = -8.49$, $p < .001$, $r = .63$, and sensemaking performance, $t(124) = -19.2$, $p < .001$, $r = .87$. We can therefore be confident about concluding that the game had an educational benefit (Hypothesis 1). The correlation between these two main outcomes had a consistent value before and after and turned out to be moderate, $r = .40$, $p < .001$. This confirms the existing literature on the care that needs to be taken with self-assessment and the need for including more comprehensive examinations (Hypothesis 2).

Much surprisingly, no differences were found in performance between virtual and real pictures, before the training as well as after the training (Hypothesis 3). This indicates a strong transfer effect. Players are capable of applying what they have learned from the game into another context. Although how participants appreciated the game does correlate with knowledge perception and sensemaking performance, it turns out that this is entirely mediated by how well they played the game (Hypothesis 4). Therefore, performance in the game explains what they gained from it as well as how they appreciated it. With this in mind, it should not be a surprise that the (average) game score correlates strongly with post-knowledge perception, $r = .54$, $p < .001$, and with the post-sensemaking test, $r = .71$, $p < .001$ (Hypothesis 7).

In terms of achieving a better performance, practice does make perfect (Hypothesis 6). The number of exercises correlates to post-knowledge perception, $r = .31$, $p = .001$; and to the test, $r = .50$, $p < .001$. Further investigation revealed that all exercises helped in the learning process. The amount of exercises was not excessive but necessary.

As for the vocabulary, the word use associated with the game almost tripled, from 9% to 26% (Hypothesis 5). Knowing that consistently 29% of the words participants used were frequent words, such as "the, is, from, that, and," it becomes clear that a large part of participants' vocabulary draws from the same source, suggesting that communication improves.

Game scores are an important factor and it turns out that computer skills moderate these, $r = .57$, $p < .001$ (Hypothesis 8). This means that participants without these skills are at a disadvantage. They will not profit from these digital tools as others. The same conclusion cannot be made about game skills. It turns out that its relationship with the game score is mediated by computer skills. However, it needs to be kept in mind that few participants had any experience in games. In future research, with more diversity among participants in game skills it may turn out to be a factor of influence.

Not only skills but also people's attitude matters. Game attitudes moderate the game judgments, $r = .48$, $p < .001$ (Hypothesis 9). This means that what people think of games is an antecedent of how they appreciate it. This appreciation, in turn, is related to knowledge perception and sensemaking performance–but mediated by game performance. Here we see that a complex network of factors plays a role in the game's effectiveness. Adding to this complexity is age (Hypothesis 10). Age has a strong, independent correlation with game performance, $r = -.57$, $p < .001$. This suggests that there is something else other than the skills to control the game in why older people gain less from the game.

5.2 Substantiated understanding

Investigation of how players experienced the game revealed that they had to learn how to "read" the game. It took some practice to learn the controls and get used to the virtual

environment. At first they were highly critical of it, rejecting it compared to the reality they know. After acclimatizing to this new environment, participants found it more fun, realistic and educational. The appreciation reached its peak at the fourth home exercise. It seems most likely that the last two exercises were too difficult. They were put in an entirely new environment with more (severe) failures.

This investigation also revealed that players were quite frustrated while playing the game. Some were frustrated about the feedback they received. The computerized Action Center with whom they have to communicate with can be sarcastic at times. Players took the game seriously and disliked how this Action Center responded:

During the start-meeting the Action Center already bothered me. I do not think we will become friends (Participant #49).

Another, very prevalent frustration concerned the “walking mouse pointer.” Due to a bug in the game engine, the mouse pointer automatically crawls upward if the mouse is not used. This seems a minor annoyance, but considering the overwhelming amount of comments made on this issue throughout the exercises it had a major impact on the experience:

The “drifting” cursor ruins a large part of the pleasure (Participant #135).

From this we can learn that a game is a sensitive medium. Small things could disrupt or frustrate a player. We can also learn that players engage in much meta-cognitive thinking by relating the gameplay experience to reality, their own standards, or their learning process. Although some players were clearly distrustful of anything virtual, many players seemed to have a healthy dose of criticism toward what is represented in the game:

If it rains and you walk over a levee, you always see puddles. Here you see one and that is immediately a failure (Participant #68).

Whether the participants are right or not (e.g., Participant #68 is not correct in his assessment), it does show that playing game facilitates the participants in thinking about their practice. We can conclude from this that the game is enabling them to become *reflective practitioners* (Schon, 1983), which are practitioners who have the capacity to reflect on action so as to engage in a process of continuous learning.

Exploration into how players actually played the game taught us that the game maybe only reached half of its potential. Participants may have learned from some of the feedback, but generally they became better in playing the game and not in making sense of virtual risks. It turns out that players stick to what they first decided on when encountering a failure. They remain consistent on how they make sense of failures throughout the game. However, their performance was mostly good from the start, therefore not necessitating significant changes. From this we can conclude that how players assessed the failures no strong conceivable learning process is visible.

This outcome is in stark contrast with the empirical evidence. To reconcile this difference, it could be that players may have learned from the mere exposure to the material. Another possibility is that being better at playing the game may have required stronger mental models of failures. Whatever the explanation, this lack of a visible learning process in making sense of failures in the game tells us that in terms of effective design, more attention is needed. What players do in the game is crucial for its impact.

Other insights from how people play is that they take it very serious. Participants really wanted to do well and were precise in making their reports and anything else that they were required to do in the game. Interestingly, in critical situations players became pragmatic, ensuring that at least a flooding was prevented before filling out the details of the report. Other observations highlight that heuristics and cognitive biases play a role in how people play. The context (i.e., critical or not?) had an influence on what players reported and over time player expectations drove what they reported. This tells us that in designing games we

have to be cognizant of human error in decision-making and provice mechanics to counter this behavior.

6. Conclusion

The game-based training with *Levee Patroller* provides for clear evidence that games can be used to help practitioners make sense of risks. The evidence also suggests that it has an impact on the communication between practitioners. These positive results confirm that games are a potentially powerful tool to enable players to make sense of phenomena. In terms of sensemaking, one can see potential in gaming sense of the rare and hard to experience (e.g., *Levee Patroller*); gaming sense of the hard to see (e.g., physics); and gaming sense of the future.

However, exploration into how players actually played the game provided information that the full potential has not been reached. Although some of the problems are intrinsically part of the learning process of using games as a training tool, other issues could have been prevented or improved. Therefore, much work can be done to increase its potential, such as including better feedback and scaffolding techniques. In addition, we need to achieve a better understanding of how players experience playing games such as *Levee Patroller* and this can be done by continuing to study what happens during playing and making use of game data (Seif El-Nasr, Drachen, & Canossa, 2013).

This study is an example of how achieving a “thicker description” by employing various qualitative and quantitative methods helps to realize a comprehensive examination of the object of investigation. Not all possibilities within the data have even been explored, but what the data does provide is that playing games is a complex activity involving many (contextual) factors and therefore deserves to be investigated in a comprehensive manner. It also shows that data provided by one source needs to be re-interpreted by data from another source. By continuing to aim for these thicker descriptions we will be able to take game research to another next level.

Acknowledgments

This chapter is based on the dissertation *Making Sense of Virtual Risks: A Quasi-Experimental Investigation into Game-Based Training*, published by IOS Press in 2012. Further details are described in this dissertation. For the completion of the dissertation and research I would like to thank Deltares, and in particular the Deltares Game Team, for their support throughout the entire study. My gratitude extends to the participating water authorities and, of course, the patrollers who were willing to participate in this study. Special thanks go out to my advisors Prof.mr.dr. Hans de Bruijn and Dr. Igor Mayer.

References

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

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago, IL: Rand McNally.

Creswell, J. W., & Clark, V. L. P. (2007). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage.

Dervin, B., Foreman-Wernet, L., & Lauterbach, E. (2003). *Sense-making methodology reader: selected writings of Brenda Dervin*. Cresskil, NJ: Hampton Press.

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

Geertz, C. (1973). *The Interpretation of Cultures: Selected Essays*. New York, NY: Basic Books.

Harteveld, C. (2011). *Triadic game design: Balancing reality, meaning and play*. London, UK: Springer.
Harteveld, C. (2012). *Making sense of virtual risks: A quasi-experimental investigation into game-based training*. Amsterdam, the Netherlands: IOS Press.
Hays, R. T. (2005). *The effectiveness of instructional games: A literature review and discussion* (No. 2005–004). Orlando, FL: Naval Air Warfare Center Training Systems Division.
Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1), 43–52. doi:10.1207/S15326985EP3801_6
Meister, J. (Ed.). (2002). *Pillars of e-learning success*. New York, NY: Corporate University Exchange.
O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: evidence from the computer games literature. *The Curriculum Journal*, *16*(4), 455–474. doi:10.1080/09585170500384529
Plass, J. L., Moreno, R., & Brünken, R. (Eds.). (2010). *Cognitive Load Theory*. New York, NY: Cambridge University Press.
Prensky, M. (2001). *Digital Game-Based Learning*. New York, NY: McGraw-Hill.
Rijkswaterstaat. (2011). *Water management in the Netherlands*. The Hague, the Netherlands: Rijkswaterstaat.
Schon, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.
Seif El-Nasr, M., Drachen, A., & Canossa, A. (Eds.). (2013). *Game analytics: Maximizing the Value of Player Data*. London, UK: Springer.
Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, *64*(2), 489–528. doi:10.1111/j.1744-6570.2011.01190.x
Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-Assessment of Knowledge: A Cognitive Learning or Affective Measure? *Academy of Management Learning & Education*, *9*(2), 169–191.
Tobias, S., & Fletcher, J. D. (2012). Reflections on "A Review of Trends in Serious Gaming." *Review of Educational Research*, *82*(2), 233–237. doi:10.3102/0034654312450190
Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.
Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., … Yukhymenko, M. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research*, *82*(1), 61–89. doi:10.3102/0034654312436980

CV

Dr. Casper Harteveld is an Assistant Professor in the Game Design Program at Northeastern University. His research focuses on i) the use of games to learn about decision-making and ii) educating people in making better decisions through games. He applies gaming in particular in the context of resilience and sustainability, and pays thereby special attention to the role of collaboration. For his work he received the Young Talent Prize of Information Systems, the Special Prize for Best Dissertation in Simulation & Gaming, and two National Science Foundation awards.