DELFT UNIVERSITY OF TECHNOLOGY, ACCENTURE, ROYAL DUTCH SHELL NV

Quasi-Experimental Evaluation Method for a Serious Game's Learning Efficacy

Wolff, Arthur Johan

4/9/2013

Committee:

Chair: Prof. Dr. Ir. Alexander Verbraeck

First Supervisor: Dr. Igor Mayer

Second Supervisor: Dr. Ir. Rens Kortmann

External Supervisor (Accenture): Dr. Ivo Wenzler

Email: arthurjohanwolff@gmail.com

Studentnummer: 1267051

"It's-a me, Mario"

(Super Mario, June 26, 1996)

Voor Margreet, hiermee los ik mijn belofte in.

Preface

Completion. An end to a very interesting time, an interval in time where the shaping of my mind, body and spirit accelerated. This process will continue in the future, but the University's environment was of a great contribution. I'm very thankful for the academic opportunity and all people that surrounded me during my time in Delft.

Summary

Background: The increase in elaborateness of serious games has resulted in a more and more prominent use in business environments, their use has grown dramatically. Due to this growing interest; Shell has requested, in 2009, the development of the serious game Hazard Recognition, a single player 3D simulation of a Shell drilling site.

Motive: The development of Hazard Recognition has reached completion, but due to the absence of experimental evidence that Hazard Recognition contributes to learning, adoption into Shell lacks. In order to collect this evidence, an evaluation study has been proposed.

Complication: Mayer et al. (2012) are in the midst of development of an overarching framework for the evaluation of serious games. However, there is no off-the shelve method available for the evaluation of a serious game's learning efficacy. Therefore, this study's research question is:

> *"What is an appropriate method, which can operate in the Comprehensive Evaluation Framework (Mayer et al., 2012), for the evaluation of a serious game's learning efficacy and the contributing factors?"*

Method: Literature research has provided for the necessary academic foundations for the development of the evaluation method. Successively, a quasi-experimental experiment shaped the context of application and testing of this method. The experiment was the evaluation of Hazard Recognition's learning efficacy. The result of this experiment showed that Hazard Recognition achieves learning efficacy.

Results: An appropriate conceptual method for the evaluation of a serious game's learning efficacy, within the context of Mayer's Comprehensive Evaluation Framework, has been laid down. The learning efficacy of a serious game is defined as the ability of the learning intervention to achieve its learning objectives. This can be mapped and measured by evaluating the movement of students through Burch's Competence Model (1970s). This requires to map the position of the student on the two axes (conscious and competence) by the measurement of self-perceived competences (conscious) and objective performance (competence). For these measurements there has been formulated an experiment setup with various requirements and constructs for the measurement methods have been developed. The case study, which consisted of the evaluation of Hazard Recognition's learning efficacy, showed that both the developed conceptual model, as the developed measurement methods are valid.

Table of Content

This page is left blank intentionally

1 Introduction

1.1 Problem Exploration

Within the oil & gas community Shell has a great track record when it comes to casualty and accident prevention (Shell, 2013a; Shell, 2013b). As a result of this constant quest for better safety, Shell has laid down a request with Delft University of Technology in 2009 for the development of the serious game simulation Hazard Recognition, adding a new literacy to their existing set of training methods. Literature indicates that using different literacies in learning the same topic improves the learning outcome (Leutner, 1993).

The development process was in close cooperation with Shell safety professionals and was finished in 2012. During the development process the learning objectives of the simulation were clearly defined; *"to identify, assess and manage site hazards effectively"* (Shell, 2009), however it is, up to this day, unclear in which extent the simulation achieves those objectives. Hence, Shell has the access to an, in potential valuable, serious game but there is no knowledge on the learning efficacy of this application. Due to this, people within Shell are reluctant towards the adoption of Hazard Recognition into Shell.

In order to remove this reluctant attitude, the simulation has to prove its value before an adoption process can initiate. Therefore there has to be gathered knowledge on the learning efficacy of Hazard Recognition, for this reason an evaluation study has been proposed to understand the potential learning benefits Hazard Recognition. This research will be committed to the development and construction of an appropriate method for the evaluation of a serious game's learning efficacy and will apply this method on Hazard Recognition in the form of a case study. The result will be twofold; a method for the evaluation of serious games' learning efficacy, and knowledge about the learning efficacy of Hazard Recognition.

1.2 Opportunity Framework Serious Gaming in Shell

This paragraph is a summary of, and based on, the Grounding Presentation Document (I. S. Mayer et al., 2009), which has been written by the developers of Hazard Recognition.

The development of Hazard Recognition commenced with the development of a roadmap. This roadmap started from the assumption that there are numerous opportunities within Shell where the application of serious gaming could add significant business value. The assumed opportunities range from operational training, emergency training to site planning and decision making. After analyzing the opportunities and weighing them against their potential value impact, Health Safety Environment (HSE) was identified as one of the most promising to commence with. This process and the topic of Serious Gaming was accompanied by lively discussion within Shell. The Delft University of Technology researcher noted that their meetings with people from Shell were full of anecdotes of how computer games are pervading social life and what this could imply for Shell. Hence, besides that Shell had little experience with serious games at that time, this shaped the perfect environment to commence with the development of a serious game simulation on a HSE topic, especially in the light of the Baker report, Goal Zero Project and the external pressure of NGO's on HSE issues, at that time.

The main purpose of the development of Hazard Recognition was to demonstrate and enhance the technical and conceptual possibilities of virtual learning and –training, using state-of-the-art commercially available game technology. The current (and final) version of Hazard Recognition aims to show the powerful potential of serious gaming for HSE. The simulation can be developed into a full-fledged application that can be implemented and validated as a game-based virtual training tool.

Thus, Hazard Recognition takes advantage of ongoing academic– and technical innovations in order to develop a proof of concept for serious gaming opportunities within Shell, and simultaneously add potentially significant business value to HSE topics in Shell.

1.3 Hazard Recognition

Hazard Recognition offers a new way of experiencing a professional environment from your desk, an experience full of interaction and direct feedback creating meaningful interactions that could improve learning (Howland, 1999). Within Hazard Recognition's virtual 3D environment (screenshots presented in Figure 1 and Figure 2) multiple real life procedures and processes are simulated, in companion with common hazards and procedural flaws where professionals in the oil & gas community are confronted with on a daily basis.

Hazard Recognition can be classified as serious game with its *modus operandi* being a single player virtual 3D simulation. The only multi-actor interaction in the game is conducted with in-characters or the during the introduction/wrap-up outside the simulation, but within the workshop, which are held in a classroom setting.

Figure 1 Through the Eyes of the Player

Figure 2 In-Game Overview 2 of Hazard Recognition

Hazard Recognition has the game play of a classic First Person Shooter. The player sees the virtual reality through the eyes of the avatar that is in control of the work that is being conducted on the worksite and the movement is operated by mouse and keyboard. Therefore, anybody who has ever played a PC video game will be able to move around in the virtual environment and start exploring the site without any necessary explanation.

Once the avatar moves around on the drilling site, the crosshair (located in the middle of the player's sight) is the interface which can activate items or characters. By aiming the crosshair on in-game characters and clicking on them an interaction window appears, allowing the player to ask preselected questions. These questions are typical procedural checks and information gathering on, for example, the quality of the equipment.

Other interaction takes place when the player detects a hazard. Once this happens, the player strikes spacebar and an extensive menu appears with possible hazards. By isolating the correct hazard in the menu, the player has successfully assessed and controlled the hazard, if the wrong procedure is chosen the hazard will ultimately fire.

1.4 Peer Simulations

There are not many serious game simulations with a combination of the same topic, objective and industry of operation as Hazard Recognition. Many virtual reality training methods are aimed to train one or more skill(s) other than detecting hazards. In surgery there are many virtual reality training applications available aiming at the practicing of a certain surgery routine or the use of equipment, a search entry on scholar.google.com on *"surgery AND virtual AND reality"* yields over ninety thousand results. A similar training objective but in a more similar environment as Hazard Recognition can be found in simulations as the Firefighting Training Simulator developed for the

US Navy (Tate, Sibert, & King, 1997), simulating the USS Shadwell, the Navy's full-scale fire research and test ship.

Virtual Rig Visit is one of the reference points of Hazard Recognition and it was developed for Shell by an external company in the late 1990s (I. S. Mayer et al., 2009). In 1999 the development of the safety training simulation for refineries *omVR* added a second reference line for simulations in the oil and gas industry (Haller, Kurka, Volkert, & Wagner, 2001). This simulation uses virtual reality glasses in order to achieve a higher degree of immersion compared to computer screen interaction and provides an improved technique for personnel safety training and it allows users to navigate through the refinery scenes and to interact with the environment. More recently in 2012, the Houston Advanced Research Center (HARC) developed a virtual rig with the purpose of familiarizing outsiders and insiders of the oil and gas industry with environmentally sustainable technologies in the energy industry. The simulation's environment is very similar to that of Hazard Recognition, with the difference that this environment is static. Players are only able to walk on the plant, discover its structures and interact with movies, educating on a specific topic (HARC & Haut, n.d.)

In the late 1990s, the South African mining industry experimented with, as they called it, "virtual reality training" (Squelch, 2001). This simulation is very similar in its objectives and gameplay as Levee Patroller (C. Harteveld, Guimaraes, Mayer, & Bidarra, 2009; Casper Harteveld, Guimaraes, Mayer, Bidarra, & Guimarães, 2007; Schuurink, Houtkamp, & Toet, 2008), but shares the industry characteristics of the oil and gas industry. Hazards in mines have to be detected and assessed by the player. An evaluation study to the success and implementation effort of an updated version of this simulation, *Look, Stop and Fix*, was conducted in 2009 (van Wyk & de Villiers, 2009) which had positive outcomes regarding the learning effect on players and the adoptability of this literacy by players.

The concept of hazard detection and assessment has been used more frequently in different industries with the use of similar simulations. Levee Patroller shares such similarities. Levee Patroller is developed by Deltares in cooperation with Delft University of Technology and aims to train levee inspectors in a virtual environment (C. Harteveld et al., 2009; Casper Harteveld et al., 2007; Schuurink et al., 2008). Despite of the different industries, Levee Patroller and Hazard Recognition share many similarities. Both simulations have the objective to train its players in the detection of system failure. However, Hazard Recognition also deals with dynamics of in-game characters and their behavior, whereas Levee Patroller is more about static failures of levees. A second important difference is the fact that levee failure can be noticed in advance before a direct danger occurs. If deviations are detected on time, the hazard can be contained. In the oil and gas industry, hazards can cause a fatality in a matter of seconds if there is no direct intervention. A simulation similar to Levee Patroller for the aviation maintenance industry was developed and tested in 2001 (Vora et al., 2002). This simulation aimed at detecting corrosion, cracks and damaged conduits on airframes (the core of an airplane).

The wide variety of simulation application in the field of hazard detection skill training is not surprising, regarding the fact that there is a widespread acknowledgement on the potential benefits that serious games and simulations can offer. Hazard Recognition cannot be regarded as a novelty in either its technique, subject or environment, but there are very little simulations that combine all three in one package. Hence, Hazard Recognition is a unique combination of environment, technique, dynamics and topic that this requires further investigation, rather than that its merits can be assumed based on peer simulations.

1.5 Serious Game Evaluation Framework

The history of games in a professional environment can be traced back nearly 5,000 years to the development of board games and war games (Faria, Hutchinson, Wellington, & Gold, 2008). Since then, much has evolved from simple chess-like games towards the current widespread availability of complex multi-actor games, 3d games and the use of elaborate in game communication features.

1.5.1 History of Evaluation Studies

As a result of the increased elaborateness of serious games, their use has become more and more prominent in business environments. As noted by Wolfe (1993); over the 40-year life of *Simulation & Gaming*, the use of business games has grown dramatically and has reached a point of relative saturation. This rapid growth in the use of serious games in business environments can be explained by the fact that many companies have become equally fascinated by the possibilities of serious gaming for learning and training, mainly with the objective to increase the effectiveness and/or to reduce the costs of learning and training (Warmelink, Meijer, Mayer, & Verbraeck, 2009). It is therefore significant to witness that, while expectations of games are; to increase learning and reduce costs, there have been few attempts to introduce frameworks that can help support tutors to evaluate games that can be most effective in their particular learning context including their specific subject areas (Amory & Seagram, 2003) (Kirriemuir & McFarlane, 2004).

The serious games that have been evaluated, have been evaluated by the use of a framework and method specifically made for that particular project. This means that each evaluation is conducted on different criteria with the use of different performance parameters. An import reason for this recurring approach of serious game evaluators is that there is very little consensus amongst serious game developers and researchers on how to evaluate the merits of a serious game (Spek, Wouters, & Oostendorp, 2011), this might originate in the disciplinary confusion and fragmentation in the serious gaming community (Mayer, Bekebrede, Harteveld, Lo, Ruijven, Warmelink, & Zhou, 2012). Consequently there was until recently the lack of an overarching evaluation framework, this made it hard to measure if the serious game introduced to a company actually met the expectations of learning improvements. However, the serious gaming community is well aware of the necessity of evidence based claims (Mayer, Bekebrede, Harteveld, Lo, Ruijven, Warmelink, & Zhou, 2012), which makes the absence of an overarching framework striking.

Consequence of the lack of an overarching evaluation framework is twofold. Firstly, there is the problem of incomparability. This means that evaluation results are incomparable between games due to the lack of an overarching framework, which defines the concepts of study and methodology. Secondly, there is a complication due to the absence of an overarching framework; there is no of the shelf evaluation framework, nor method, for newly developed games.

1.5.2 New Evaluation Development

Mayer et al. (2012) are in the midst of the development of an overarching framework (Figure 3), such a framework would be capable of overcoming previously mentioned complications. This research will use this framework as the starting point for development of the evaluation method and the design of the experiment. Because this framework is in the midst of its maturing process, this research will be the first application of the framework. As a result hereof, the merits of this study will not only relate to the evaluation of Hazard Recognition, but also consist of valuable knowledge on the application of the framework that can be used by the developer for further improvement. This knowledge will consist of methodological application development within the framework and possibly rethinking certain methodological aspects for the use on serious games. Hence, this research will contribute to the academic merits of Mayer et al. (2012) by building further on the foundation in order to construct an evaluation method for the learning efficacy of serious games and by testing this method on the serious game Hazard Recognition.

Dividing the Framework

The top level of the framework defines the influence of the game on the total system. Such influences can only be measured by evaluating the total system, in this case the whole Shell enterprise, and follow its performance over the years. This will not be the focus of this study, because the timeframe doesn't facilitate this. The focus of this study will be on the *Individual and team level in the simulated game environment*, i.e. the direct effects of the serious game on the participants is evaluated.

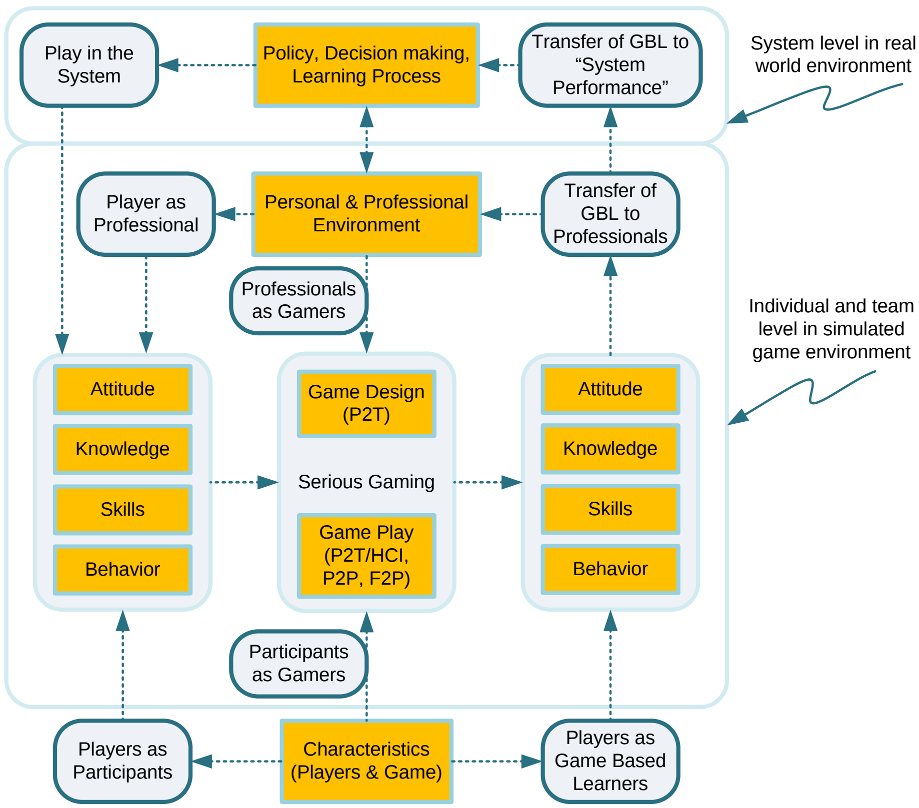

Figure 3 Evaluation Framework Mayer et al. (2012)

1.5.3 Relation Framework and Methods

The measurement methods that are used for evaluating serious games, are closely related to the evaluation framework in which they operate. The framework defines *what* aspects are measured and analyzed, and *why* they are measured and analyzed. The measurement methods define *how* this is done. This is significant to take into consideration when researching literature for methods to use within the framework. Some methods might be very applicable in the designed experimental frame, but are not relevant in the used framework.

1.6 Knowledge Gap

The problem exploration confronts us with a knowledge gap that is twofold:

First, there is no off the shelve evaluation method available for the evaluation of a serious game's learning efficacy. The Comprehensive Evaluation Framework of Mayer et el. (2012) is available, but this does not incorporate an off the shelve learning efficacy evaluation method.

Second, there is no knowledge about the learning efficacy of Hazard Recognition and its ability to achieve the learning objectives that have been set.

1.7 Study's Relevance

The proposed study is relevant in both the academic world and in society, because of the academic contribution to an existing societal issue; this section will elaborate further on this.

1.7.1 Academic relevance

The evaluation of serious games is an immature field of research, there are many leaps possible in this field. A recent leap is the development of an overarching evaluation framework by Mayer et al. (2012). Building further on the merits of Mayer by developing a method for evaluation of serious game's learning efficacy and the application hereof on an existing, but yet unevaluated, serious game would contribute to this development process and would help validating the framework's value.

1.7.2 Social relevance

The social relevance lies in the assessment of the serious game Hazard Recognition. If this training approach appears to be successful in the teaching of recognizing hazardous situations on construction sites, this means that Shell employees can be trained better and more cost efficiently. Better training will likely result in less casualties and cost efficiency will result in a financial surplus for the company, both are relevant for society.

1.8 Research Problem

The research problem is shaped by two components; the absence of an off the shelve method for the evaluation of a serious game's learning efficacy, and the absence of knowledge on the learning efficacy of Hazard Recognition. The research question should be defined and answered in the space shaped by the two components. In addition, since the Mayer's evaluation framework is the most extensive that has been found in literature, the evaluation should take place within its boundaries. Therefore the research questions for this study is:

> *"What is an appropriate method, which can operate in the Comprehensive Evaluation Framework (Mayer et al., 2012), for the evaluation of a serious game's learning efficacy and the contributing factors?"*

Sub-Questions

The sub-questions decompose the steps that are required to answer the main research question. The questions investigate the serious game and work towards the development of evaluation methods within the given framework. These methods will answer the question whether Hazard Recognition delivers any learning efficacy and consequently the game's adoption power into Shell will be investigated.

1. *"What is a conceptual model for a serious game's learning efficacy and how can this be achieved by a learning intervention?"*
2. *"How can the learning efficacy, and underlying factors, of a serious game be measured?"*
3. *"How should the evaluation framework be operationalized for the evaluation of Hazard Recognition's learning efficacy?"*
4. *"Does Hazard Recognition achieves learning efficacy and what factors contribute hereto?"*

2 Learning Concepts

Learning is the first thing man does at birth and will possibly do until his last breath. People learn every day, in every possible environment and on various subjects. Man walks the walk, talks the talk, but might also know something about linear algebra. Walking and talking might be a skill considered as something natural, but is very well a skill acquired by learning and training. Serious games can function as learning interventions and will consequently have learning objectives. The extent in which a learning intervention is able to achieve its learning objectives can be called the serious game's learning efficacy. In this chapter there will be investigated when learning efficacy is achieved and what conceptual models can help to identify learning efficacy. Therefore this chapters research questions is:

> *"What is a conceptual model for a serious game's learning efficacy and how can this be achieved by a learning intervention?"*

In order to reach the answer to this question there will be made use of different models derived from literature and consequently they will be combined in order to create a foundation from where a further operationalization and construction of measurement methods can take place in successive chapters.

2.1 Competence Development

The development of competences is done by learning and training. Ultimately trained competences can reach a state of improvement saturation, a state where improvement is no longer really achievable but the skills or knowledge should be maintained. For example driving a car, on some moment in time a person can drive a car perfectly as a commuter and is confident doing so. In this case the saturation point is achieved. When that same person decides to compete in stock car races, the competence benchmark increases and there is more room for learning.

The stages a student goes through when acquiring a competence has been conceptualized by Noel Burch (Burch, 1970s), this model is shown in Figure 4 and is important for the assessment of a serious game's learning efficacy for two reasons;

First the model does not only takes into consideration the improvement of a student's objective measurable competences but also the cognitive change that accompanies the process of learning.

When an assessment method can map the ability of an serious game in which extent it can facilitate this change in conscious, the method is able to value the serious game's ability to create awareness with the student. In other words; learning is not only improving a competence but is also creating awareness of ones (in)ability to perform a task.

Second, as a result of this multi-dimensional assessment it is possible to understand were the most potential of a serious game lies; learning competences, or creating awareness. This understanding assists the decision making process in which part of a training the serious game is best applied.

The Competence Model is relevant for competences whether they are skill based or knowledge based. According to the Competence Model the learning process starts in the quadrant where a student is typically unconscious of his incompetence. A striking example is people who are totally unfamiliar with ice skating, overestimate their competence and fall almost directly after their first try on the ice. Hubris is the exponent of behavior witnessed amongst these students.

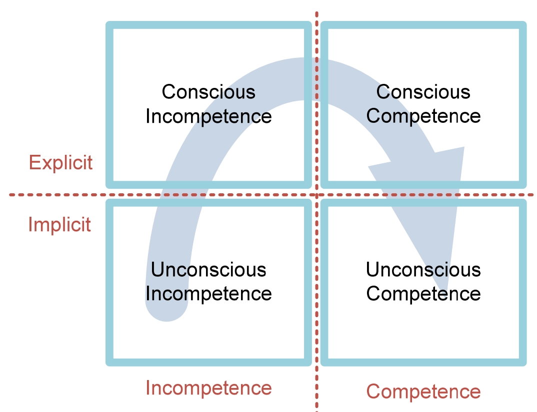

Figure 4 Competence Model (Noel Burch)

The process proceeds into the state where the student becomes conscious of the incompetence. This state is comparable to the *valley of despair*, a term commonly used in learning theory; describing the frustration of the conscious that one is not as competent as one initially expected. The transition between the first two stages does not necessarily include competence improvement, but is rather a process of becoming self-aware.

When the learning process further continuous, the student starts to improve the competence and is conscious about the recently acquired competence. Once a student is familiarized with the competence through training and application in real life, the competence is being nested. Once this nesting process is sufficient, the student is unconscious of the acquired competence and will be applying it naturally.

2.2 Definition Learning Efficacy

There are multiple ways to define learning efficacy, because of this multitude of interpretations it is recommendable to agree upon a definition that will be used in this study. Garris defined a basic distinction between skills based learning outcomes, cognitive outcomes and affective outcomes (Garris, Ahlers, & Driskell, 2002). Learning efficacy for this research will be defined as the ability of a serious game to move the student in the right direction through the quadrants of Burch's Competence Model. The competences that are the scope of this improvement depend on the learning objectives of the specific learning model. The ultimate goal of a learning intervention, in this case a serious game, is to facilitate the student to ultimately become *unconsciously competent* in the objectified competence.

2.3 Learning Interventions and Feedback Dynamics

Learning models are modes of intervention aiming on improving the student's cognitive ability, such as skills, knowledge, behavior, etcetera. Learning models come in many different shapes and forms. Regarding the scope of this research, serious games are learning interventions with learning models incorporated in them, within the a game or simulation smaller learning interventions are embedded. Hence, the bigger learning model with one specific objective is build-up of multiple smaller learning models containing their own specific objectives and modes of intervention. In order to understand the different forms of learning interventions and feedback dynamics, it is important to first understand the distinction between learning intervention's implicitness/explicitness.

2.3.1 Implicit and Explicit Learning

Learning models can be divided into explicit learning models and implicit learning models. The general definition for both learning models is: explicit learning is the acquisition of knowledge and/or skills whilst being aware of this, implicit learning is acquiring knowledge and/or skills in an accidental manner. Transposing this definition to the "walk the walk and talk the talk" we can say that when a baby learns to walk, it does this on a trial and error basis, not being fully aware of the fact he is learning something. This is implicit learning. When a grown man in a recovery center learns how to walk, this works on an explicit learning basis.

The distinction between implicit learning and explicit learning models is important for the understanding of the learning concepts imbedded into a serious game. When a learning theory prescribes an implicit or explicit learning method, this has implication on the type of learning intervention method and the type of feedback dynamics that accompanies this.

2.3.2 Feedback Dynamics

Serious game simulations create a constant two way flow of information between computer and player. Depending on the simulation and the components that are incorporated in it, there can be found different kinds of feedback in a simulation. Modern theories of effective learning suggest that learning is most effective when it is active, experimental and problem based, this is achieved through the use of, and the variation with, different kinds of feedback (Boyle, Connolly, & Hainey, 2011; Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). In a serious game simulation, such feedback serves a purpose that can be derived from the game's objective and/or learning model. In some cases this can be corrective feedback, but feedback can also serve clarity about content or can be of a more instructional nature. In any case, feedback can differ in its dynamic character. Feedback with high adaptability to the scenario's generated by the player is dynamic. On the contrary, static feedback is not adaptive and is always available to the player in the same form.

In addition, the design choices that are made about the explicitness of the learning interventions have important implications on the dynamics of the feedback and the learning outcomes (Leutner, 1993; Moreno & Mayer, 2005). Corrective feedback on certain actions can contain solely "correct" or "incorrect", guiding the player in the right direction by giving feedback on the performed action; if it was correct or not, this same feedback could also contain an explanation about the specific action; supplying more in depth information. The adaptability of the feedback is more or less the same. An argument could be made that the more extended feedback is also more adaptive, but certain is that the extended feedback is more explicit than the other since it operates with explicit learning material than it is operating on a trial and error base. This means that the feedback dynamics play a role in the design of the learning model.

Since the learning efficacy is achieved by the quality of the learning models, these learning models should be investigated. This results in the approach that a proper evaluation does not value the simulation merely as a black box affecting the initial conditions of the participant, the simulation is a process with dynamics that could be altered in order to achieve better learning. When the game is decomposed, each composition will be mapped on the static – dynamic scale, this will help to identify how the feedback is organized and during the evaluation this map can be useful later in the evaluation process to determine if certain information might achieve better learning when it is made more or less dynamic.

2.4 Kolb Learning Cycle

The Kolb learning cycle is an iterative cycle that indicates the preferable states that should be incorporated in a learning model in order to achieve a good functioning learning process (A. Y. Kolb & D. Kolb, 2008; D. Kolb, 1984). There has been chosen for Kolb's learning model, for the definition of a learning process, because "Kolb's Learning Style Model is the most widely accepted learning style model and has received a substantial amount of empirical support" (Manolis, Burns, Assudani, & Chinta, 2013). Kolb's Cycle relates to the definition of learning efficacy (paragraph 2.2) because the acquiring of a competence by learning/training is described by Kolb's model, consequently this learning process facilitates the movement through Burch's model.

For the evaluation of a serious game's learning efficacy one must look at the ability of the game's learning interventions to fulfill the requirements of the Kolb Cycle. If there is insight of the ability of the simulation's learning interventions to facilitate a Kolb cycle, there is also insight in the full potential of the game to train/educate players. After all, if there are no closed cycles, learning would be hindered, ineffective and/or slow.

The four states that form the cycle are successive, iterative and continuous.

1. The cycle typically starts with the student performing a task and witnessing the effects hereof, called the experience.
2. The next step, the reflective observation, would be the understanding of this causality of this particular instance so that same situations, later in time, would be anticipatable.
3. The third successive step, the abstract conceptualization, is the generalization of the underlying principle in order to develop knowledge and skills in such a way that similar situation can be tackled with success in the future.
4. The fourth step is the application of the understood principle in new situations under different circumstances, the active experimentation. With this understood principle in hand, the iteration would start with a new experiment, under different circumstances.

These new circumstances in a different environment will create the application of new principles that can be understood. This process is usually represented in a circle, but in reality it is a spiral. Because if earlier experienced principles are understood, only the introduction of new principles would require the steps of generalization and understanding in order to acquire new knowledge or skills.

All the states in the Kolb Cycle require cognitive processes of the player. These cognitive processes are described in the Interactive Cognitive Complexity (ICC) Model (Tennyson & Jorczak, 2008).

2.5 Interactive Cognitive Complexity Model

The Interactive Cognitive Complexity Model (ICC Model), shown in Figure 5 (Tennyson & Breuer, 1997, 2002), is "a conceptual framework that identifies the key variable of instructional games affecting player motivation and learning [...] The ICC Model is an integrative information-processing learning model that views learning as the result of complex and non-linear interaction of variables internal and external to the cognitive system of a learner." (Tennyson & Jorczak, 2008). This model can help game developers to effectively design games for instructional use because it gives insight and overview to the variables and cognitive processes which the game should influence in order to achieve learning efficacy. Hence, the cognitive capacity of the student, as defined in the ICC Model, is not only subject of improvement by the learning intervention but also the *modus operandi* to achieve in the game and achieve learning.

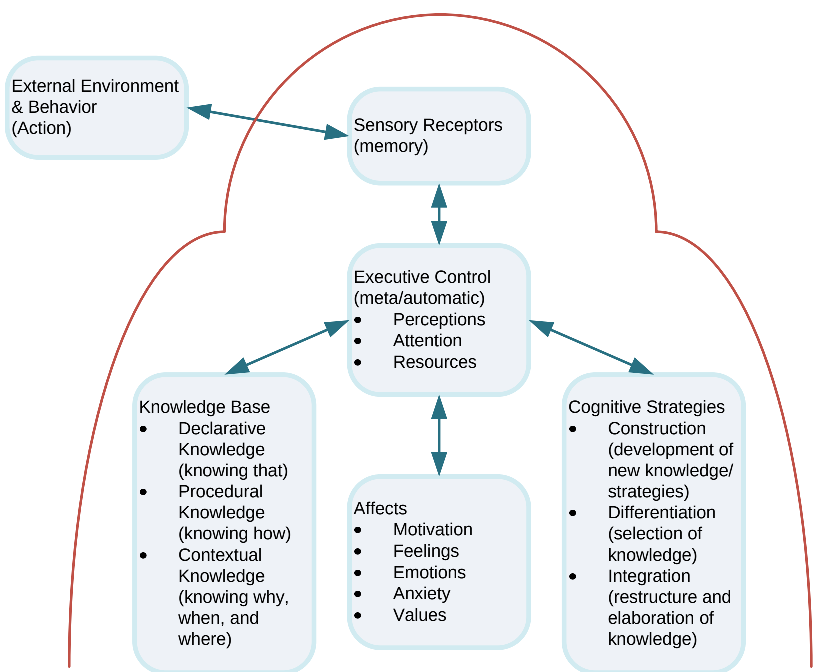

Figure 5 Interactive Cognitive Complexion Model (Tennyson & Breuer, 1997, 2002)

The cognitive processes that are defined in the ICC Model relate strongly to the learning stages from the Kolb Cycle because different stages in Kolb's cycle require different cognitive tasks, which are separated and linked in the ICC Model. Since a good learning intervention has a closed Kolb Cycle this means that the assessment on a game's learning efficacy should take place in a balanced manner on all the student's cognitive capacity in order to map the full spectrum of change the game accomplishes.

2.6 Hazard Detection According to the Kowalski Model

In previous paragraphs the conceptual model of the assessment of a serious game's learning efficacy have been explained. As a validation of this conceptual model the simulation Hazard Recognition will be assessed on its perceived learning efficacy. In this paragraph there will be presented an academic validated conceptual model that describes the process of hazard detection and will therefore function as the set of competences that Hazard Recognition aims to improve according to its learning objectives.

This conceptual model of hazard detection is developed by Kowalski (Kowalski-Trakofler & Barret, 2003), shown in Figure 6. This model can be considered as the fundament of the learning model Hazard Recognition aims to deliver to its players. Hence, the skills defined in Kowalski's model are the competences that Hazard Recognition should improve amongst its players in order to establish a part of the objectified learning efficacy.

Kowalski's model is not only conceptual but did also benefited from basic laboratory research into human information processing and studies of the development of expertise in complex tasks (Perdue, Kowalski-Trakofler, & Barret, 1995) and the model has been validated. This model distinguishes and defines the skills that should be present with the observer in order to detect hazards and is therefore very suitable for this study. It is a strong conceptual model that exactly describes the tasks and sequences Hazard Recognition trains. Three interesting remarks can be made on the model's description and deserves to be explained here before it can function as part of the evaluation process. The logic of the explanation is; Remark per process step → consequence in general for the detection of hazards → put in perspective of Hazard Recognition → consequence for the use of Hazard Recognition.

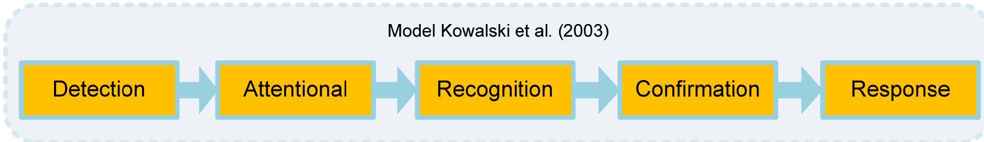

Figure 6 Hazard Detection Model Kowalski et al. (2003)

Detection: Kowalski finds that discriminating "figures" from other stimuli serving as a "ground" for the sensory of hazard cues. The consequence might be obvious, but should be made explicit; the observer is only capable in detecting deviations from this baseline if the observer has knowledge about this baseline. Putting this in perspective with Hazard Recognition, there is a reciprocity between the level of knowledge of the player; which determines his/her ability to detect the deviation, and obviousness of the hazards presented by the game. This reciprocity determines, most likely, the learning efficacy. The proportion between the games complication and the participant's ability to solve the objectified tasks is also the main determinant for *Gameflow* (Alvarez, 2008). Gameflow is the state of the player wherein (s)he is willingly enough to proceed with the game, this commitment is created by presenting challenges and rewards for the achievement of those. The best gameflow is achieved when the difficulty and reward are an optimal balance for that particular player.

Recognition: This phase is strongly funded on perceived existence of tacit knowledge with the participant. This can be, according to the model, trained or gained from experience. Consequently, in perspective of the simulation and the supporting workshop this might not always be expected due to the participation of novices in the simulation.

Confirmation: "People seem generally reluctant to disconfirm any tentative hypothesis they hold" (Watson, 1963), according to Kowalski the efficacy of this stage is determined by the ability to avoid a *confirmatory bias*. This is true, however this is also the obvious answer, promoting people to be open to reconsideration on their hypothesizes would be a good strategy to avoid biases. In light of Hazard Recognition, we can say that reconsideration is forced upon the player by the fact that wrong decisions are corrected by direct initiation of accidents. A consequence might be that this would favor a trial and error approach, rather than a deliberate decision.

2.7 Workshop Decomposition

For the proposed case study that will be used to validate the conceptual model that has been developed earlier in this chapter, it is of particular interest that the simulation is decomposed to all individual learning interventions according to the developed conceptual model. Once this has been done it will become clear if Hazard Recognition can be considered as a complete learning intervention, as it was defined with the combination of Burch, Kolb, and Kowalski.

The workshop consists of two elements, the game and the presentation. The presentation delivers its content in a consistent explicit way. The simulation however, is build up with different components, each of them differ in their interaction style, explicitness and literacy. Therefore, it is not possible to examine the simulation on one aggregation level, but it should be decomposed.

Each particular learning intervention facilitates in theory a full Kolb circle on a micro level, but on a macro level does the whole workshop (simulation and presentation) facilitates one bigger Kolb cycle. The game would achieve, in theory, the highest learning efficacy (given its learning content and scope) if each learning intervention cycle forms exactly the necessary basis of knowledge and skill through conceptualization for the participant to enter the next learning intervention. The decomposition from macro level to micro level can also be represented on the axis general → detail. The big workshop cycle has a higher level of abstractness than the lower cycles that are represented in the simulation on very specific detail of a hazardous situation. This decomposition of cycles is visualized in Figure 7.

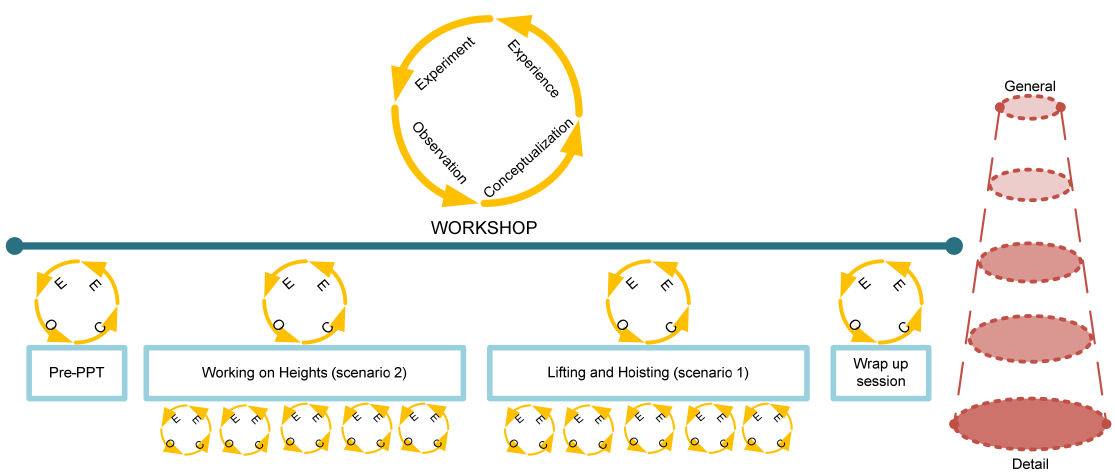

Figure 7 Workshop Decomposition of Kolb Cycles

Hazard Recognition has been developed with a clear learning objective in mind, defined in the Project Initiation Note (PIN) (Shell, 2009); *"Identify, assess and manage site hazards effectively"*, and has been developed as an instructional simulation game with learning outcomes as primary goals (Garris et al., 2002). Because the game has the primary goal to deliver learning to its users, we may expect clear, predefined learning models. This is not the case , the game was developed with a pure focus on the realistic simulation of procedures and accompanying hazards. Therefore the learning lies within the use of the simulation, rather than that the simulation serves a predefined learning model. The purpose of the simulation was to foremost create an environment that facilitates procedures and situations copied from reality with useful interaction facilitating immersiveness (Howland, 1999).

On first sight, Hazard Recognition looks like an ordinary simulation of the Simwell facility. This is true for the environment situated in the simulation, but there is more to it. The developers incorporated in this static environment dynamic scenarios of situations one might encounter in real life at the Simwell facility. The goal of the simulation is to learn players about the detections of an intervention on hazards on worksites. Therefore, these dynamic scenarios create multiple hazards and dangerous situations. Together they form the main literacy: the experience of the Simwell facility and the encounter with real world dangers in a virtual simulation gaming environment. The simulation's content can be further decomposed into 6 ways of interaction between the player and simulation, allowing to transmit information that can be transformed in knowledge and skill. This transformation between sensory and information processing is a cognitive process described in the ICC model (Tennyson & Jorczak, 2008), elaborated on in the previous paragraph. The decomposition of the workshop is shown in Table 1 Workshop Decomposition and further explained in 0 and 2.7.2.

Workshop	
Simulation	**Presentation**
Simwell Environment	Verbal
Static Hazards	Visual
Movie Clips	Question Feedback
Conversations	
Feedback on Reports	
Procedural & Dynamic Hazards	

Table 1 Workshop Decomposition

2.7.1 Simulation

Simwell Environment

The virtual world that forms the main structure of the simulation is a static environment which embodies the main features of the real Simwell site, such as buildings and installations. The environment can be explored by the player by virtually walking around and facilitates the basic body wherein other feedback literacies operate. The interaction between player and environment is limited to the possibility of exploration by walking around in the virtual environment.

Static Hazards

Static hazards are single smaller hazards present in the virtual environment that require no procedural processes in order to observe them. They are deviations on a physical observable way, but are static in nature and do not move. The set of hazards consists, amongst others, of: dislocated fire extinguishers, gaps in the barrier tape, wrongly placed paint cans, etcetera. The interaction between player and object is hosted by the environment, once the player observes the hazard, it can report the hazard by using the reporting function. Depending on the level, the reporting is done by using the extended menu (through spacebar) or by aiming and clicking with the crosshair.

Procedural & Dynamic Hazards

The procedural & dynamic hazards are more complicated than static hazards, but in essence simulate hazards in the same way. These kind of hazards are more complex than static ones due to their nature of movement that are part of strict procedures. When the movement of objects is not conducted according to procedure or during the movement supporting elements are not working according to procedure, the hazard can fire. The detection of these hazards requires more skill, analyses and knowledge than only a gut feeling, which is sometimes enough to detect a static hazard.

Feedback on Reports

When in the simulation hazards are reported, there will be given feedback on this report by a pop up screen whether the report that has been filed is actually correct or not. There is not given any accompanying context about the report that has been filed. The interaction between player and simulation is dynamic, but limited in communicated information and is only available when the player actually intervenes on hazards. However this learning intervention may seem small, it is maybe the most important one because of the ability to directly confirm towards the player if his actions were right or wrong. This assumption is based on the findings of Leutner (Leutner, 1993), which found that system-initiated adaptive feedback can improve learning of domain knowledge .

Conversations

The conversations between player and in-game characters that are embedded in the simulations, provide not only knowledge about specific procedures, rules and conditions, but are core of the gameplay. Since the conversations are key to start certain important processes in the game. The conversations are dynamic and the interaction between player and simulation goes through a preset menu where the player can choose which questions should be asked. Those are answered by the in-game character and a new set of follow-up questions is presented to the player. These conversation scenarios are exhaustive and are limited in their content to the scope of the game and processes where the conversation is part of.

Movie Clips

The movie clips in the simulation are just as the 'direct feedback on reports' system-initiated adaptive feedback. However, the movie clips provide far more context and knowledge hazards. When hazards are not detected, the simulation will simulate the consequence of the undetected hazard. The movie clip brings a different literacy to the table during the game, this might have an extra powerful effect.

2.7.2 Presentation

The presentation consists of two literacies: verbal and visual. The content that is being covered in the presentation is to shape the context of the simulation, how the simulation works and covers some basic knowledge about hazards. The latter is significantly important for workshops where non-oil-and-gas professionals participate. The reason that this is done, is because players need at least a minimal amount of knowledge to play the game and then know what to look for.

2.7.3 Decomposition Placement in Kolb Cycle

In order to achieve the learning objective, it is essential that the, simulation its learning interventions create a closed Kolb Cycle and that the chronology of the subsequent interventions represents the same phases as the Kolb Cycle. Therefore, all learning interventions in the simulation are ranked in table 3 to their facilitation efforts of the Kolb Cycle. This brief overview shows that all four stages which are facilitated. From this overview we can conclude that all stages are facilitated in the simulation. The second requirement which should be looked into is the chronology of the interventions.

Kolb Cycle Representation				
Learning Intervention	*Experience*	*Reflective Observation*	*Abstract Conceptualization*	*Active Experimentation*
Simwell Environment	4			4
Static Hazards	4	2		4
Procedural & Dynamic Hazards	4	2		4
Feedback on Reports		4	3	
Conversations	2			2
Movie Clips		4	3	

Table 2 Learning Intervention's Kolb Cycle Representation on a scale from 1 to 4

In a typical simulation workshop, the interventions would succeed each other in the order Kolb Cycle prescribes, namely:

1) One enters the Simwell Environment
2) One is confronted with situations wherein one can more or less chose to:
    a. Start a conversation
    b. Detect a static hazard
3) After a conversation a procedure is initiated wherein:
    a. Dynamic hazards fire
    b. Static hazards fire
4) When detected on time, feedback on the report is given
5) When hazards fire, movie clips show the consequences hereof

However, in some instances players can unfortunately skip hazardous situations without knowing and are then afterwards faced with reflectionary feedback because of the unrecognized hazards. This problem occurs due to players that are facing difficulties with understanding the gameplay or scenario's. In addition, Table 2 Learning Intervention's Kolb Cycle Representation shows that the learning interventions are more strongly facilitating the Experience and Experiment stages than the Conceptualization stage. This has two reasons. First the development of a serious game simulation is all about creating an experience on a rather realistic detail, in contrast to more abstract serious games. This results in a stronger emphasis on the creation of learning interventions which facilitate a realistic experience than there is room for interventions concentrating on the abstract conceptualization. The second reason strongly relates with the first, because there is more emphasis on the realism of things, the abstract conceptualization cannot be performed on that same level of detail. Therefore, it might be so that the player conducts this process internally without the use of the simulation.

Therefore it can be concluded that Hazard Recognition has a working learning model according to various learning theories and offers set of learning interventions and literacies that facilitate the player in the conceptualization of hazard detection.

2.8 Structural Bugs

There have been found bugs in Hazard Recognition. These bugs limit the direct learning potential of the simulation because they hinder the process, procedures and game flow. However, these bugs are not structural and do not exclude the total capability. Thus, they are minor flows, not really affecting the whole learning intervention. The bugs that were found are noted in Appendix IV.

2.9 Conclusion Learning Concepts

This chapter has been dedicated to the development of a conceptual model that can define a serious game's learning efficacy and concepts that identify requirements for serious games in order to achieve learning efficacy in the best way. Therefore the research question of this chapter is:

> *"What is a conceptual model for a serious game's learning efficacy and can this be achieved by a learning intervention?"*

For a serious game in order to achieve learning efficacy the game should facilitate the student in the process of improving/changing his or her competences (Garris et al., 2002). This process is conceptualized by Burch (1970s) in the Competence Model and takes into consideration not only the improvement of the competence but also the process of becoming (un)conscious about the acquired competence; the ultimate goal of an learning intervention is to make the student *unconscious competent* in that particular competence.

For a serious game to be successful in moving the student through the quadrants of Burch's Competence Model, the game should have a set of learning interventions that together fulfill the requirements specified in Kolb's learning cycle (A. Y. Kolb & Kolb, 2008; D. Kolb, 1984). Such a cycle facilitates the student in the process of experimentation up to conceptualization of a specific competence. An incomplete cycle will result in hindered, ineffective or slow learning.

The tasks a student has to perform in order to make a Kolb Cycle iteration requires cognitive tasks. The Interaction Cognitive Complexity Model (ICC Model) maps and relates these cognitive capacity. It can be assumed that the combination of the four steps in the Kolb Cycle require all cognitive concepts defined in the ICC Model (Garris et al., 2002; D. Kolb, 1984; Tennyson & Jorczak, 2008), this means that a balanced assessment on the cognitive concepts is able to discover the capacity of the student to achieve the learning intervention's objective. Consequently, this information is also a qualifier for the learning efficacy of the serious game.

The conceptual model that has been laid down has been used to assess Hazard Recognition. The competences that should be mapped in Burch's Model are derived from the academic validated Hazard Detection Model of Kowalski (2003). In addition, it can be concluded that Hazard Recognition, as a learning intervention, facilitates all four phases from the Kolb Cycle; making it a complete learning intervention and the learning efficacy can be measured by the assessment of student's cognitive capacities from the ICC Model.

3 Experiment Design

In the previous chapter there has been formulated a conceptual model that defines a serious game's learning efficacy and the concepts that require to be measured in order to evaluate a serious game's learning efficacy. This chapter will focus on the design of an experiment, i.e. the operationalization of the previously developed conceptual model. In addition, the designed experiment will be applied for the assessment of Hazard Recognition. Therefore, this chapter's research question is:

"How can the learning efficacy of a serious game be measured?"

In order to answer this question the following steps will be taken. First, the Comprehensive Framework will be operationalized. Second, the experimental group, the participants, will be defined. Third, the concepts that will be measured within this framework will be identified and explored. Fourth, the identified concepts will be further operationalized into variables and assigned to the respective classes. In this stage, the measurement constructs will be developed or derived from existing literature.

3.1 Operationalization of the Comprehensive Framework

The framework is divided in to two system levels: a higher aggregated organizational level and an individual level. The organizational level gives insight in the effects of a serious game on an organization, after people from within the organization have been participating in the serious game. These effects can be various, but the outcome would always aim to be beneficial for the organization. Because the individual elements are the building blocks of an organization, these elements are also taken into account in the framework. This level, the *Individual and team level in the simulated game environment* is the level where this study will focus on because the need statement and research question focus on the learning efficacy. This efficacy is achieved through the learning on an individual basis. A second reason is that individual effects can be measured in relative small time; the effect of a serious game within an organization at large would require a big time span to let the effects settle down and be noticeable.

Thus, in order to gain insight in the learning efficacy of a serious game, Hazard Recognition in specific, and the factors that explain this, the Comprehensive Framework requires operationalization. Because of the scope of this research and the limitations of the resources committed to this project, this research will be limited to the analysis of the effects of Hazard Recognition on the lower level: Individual and Team Level.

3.1.1 Performance Indicators

The framework uses an input-output model for improvement of skills, knowledge, attitude and behavior. The learning objective of Hazard Recognition is to "identify, assess and manage site hazards effectively". Thus, skills and knowledge will be operationalized and are key performance indicators for this research. Attitude towards hazards, but also towards gaming and game based learning is a possible underlying factor that could explain the learning efficacy of Hazard Recognition and will therefore be operationalized. The factor behavior will not be operationalized because the behavioral change due to participating in the simulation is something that requires a longer-term investigation on the participants. The simulation experience will, according to Kolb's Cycle, be conceptualized and the change in behavior due to this will therefore only be observable when the participant is back in its professional environment. Since the resources to this project do not support such an investigation, this factor will not be taken into account.

3.1.2 Individual Factors

The individual factors that create the participants characteristics are a core in the explanation of experiment outcomes. In the Comprehensive Framework there are five individual components that indicate a direct influence on the learning outcome of players. These components are:

1) Player as Professional,
2) Player as Participant,
3) Player as Game Based Learner,
4) Participant as Gamer,
5) Professional as Gamer.

These components show a similarity with the approach of Harteveld (Casper Harteveld, 2010), in this approach there is made use of three interacting worlds: Reality, Play and Meaning. The characteristics of the 'gamer' are embedded in the Play world, whereas the 'participant' and the 'professional' come from the Reality world. The Meaning world represents the sense making and learning domain, Game Based Learner characteristics can be found here. This triangular relationship represents the three main background components of participants that should be taken into account in order to cover the whole spectrum when investigating the underlying factors that possibly explain the learning efficacy of Hazard Recognition. What we are looking for are relationships between background factors, which we can measure, and what happens during the workshop. For example, the shift in attitude or knowledge of the player.

3.1.3 Environmental Factors

The Professional and Personal Environment is a factor of influence on the players professional components, this factor is influenced by the bigger organizational system and the Transfer of Game Base Learning to Professionals. This factor will not be taken into account in the operationalization for two reasons. First, because this factor is not directly influencing the game, but is an underlying factor of an individual component which could be an explaining factor, therefore it doesn't contribute to the answer on the research question. Second, the scope of this research is the serious game, the game's perceived learning efficacy and explaining factors that contribute to the learning efficacy. To assess whether a players professional and/or personal environment is contributing to the learning efficacy of the game, requires a grand investigation into both organization and personal environment. This is very much out of scope and therefore will not be taken into account.

3.1.4 Serious Game

The center of the framework is the Serious Gaming box, it is the main subject of this study and consists of Game Design and Game Play. Both will not be operationalized in the way that they are being measured, but will be explored on their performance of improving the KPI's. The performance of the simulation depends on the overall learning efficacy it brings to players. Therefore the exploration only serves the understanding of what exactly happens in the game and which mechanics are responsible for the learning efficacy. When that is understood, possible improvements on the game design can be proposed in order to achieve better results in the future. Part of the exploration of the learning interventions on a conceptual basis have already been conducted earlier in this chapter, this exploration will form the basis for the explanation of the derived results.

3.1.5 Operationalized Framework

Thus, in this sub chapter there has been elaborated on the operationalization of the Comprehensive Framework developed by Mayer et al (2012). This operationalized framework forms the basis for the continuation of this evaluation study and is presented in Figure 8. The highlighted items in the visualization are the concepts that will be worked out further in the coming paragraphs, the shaded concepts will not be taken into account during this evaluation study.

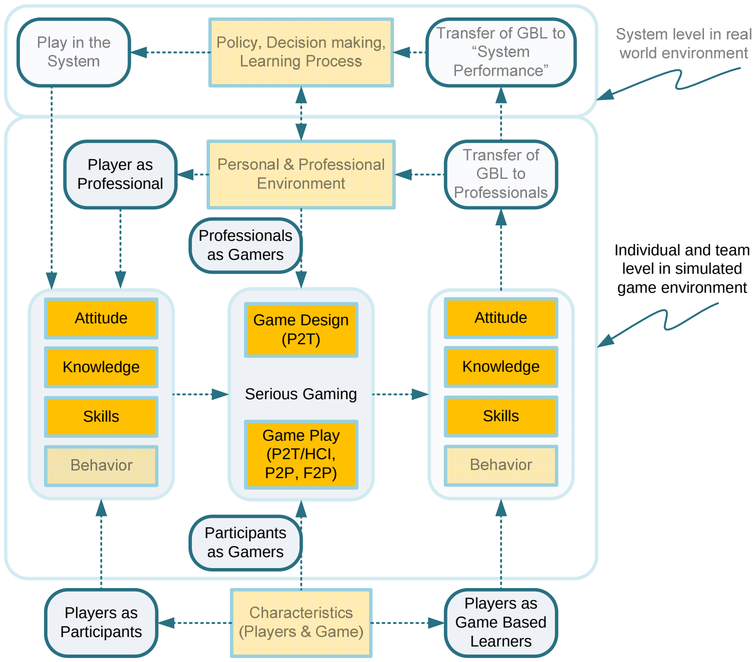

Figure 8 Operationalization of the Comprehensive Framework

3.2 Experimental Group

The experimental group is the pool of players that will participate in the workshops in order to gather data. This pool will consist of people who have a background in the oil and gas industry and people who do not. Because this research is scoped into the Shell organization, the professionals with a background in the industry will all have some sort of employers relationship with Shell and the participants from outside this industry will be affiliated with Delft University of Technology or with Accenture.

The researcher has chosen for two distinctive groups in order to create a clear distinction between a competent and an incompetent group, as defined in the Competence Model. The group with a background in this industry has extent or moderate relevant experience accompanied with relevant knowledge and skills, in contrast with participants without a background in the industry who have no relevant experience, relevant knowledge nor skills. This means that both groups have a different initial state in the Competence Model.

These characteristics are embodied in the *individual factors* and should be treated as such. However, this general distinction within the experiment group has a major influence on the workshop because of entry level knowledge which might need compensation and affects therewith the ability of measurements. In short, it cannot be expected that participants not coming from the industry have any relevant knowledge and skill that could be measured objectively; questions about procedures and factual statements. Where those participants can report on is on their self-perceived competence within the given context. This would be an perception of their own ability.

3.3 Observations in Quasi-Experiment

Following the logic of the Comprehensive Evaluation Framework, the serious game is a treatment that alters the players initial levels of knowledge and skills. In order to understand this process, and ultimately the efficacy of this treatment, the experiment should be capable in revealing the effects. Due to the limitations of the available resources and the finite amount of participants (paragraph 3.2), it is impossible to have a randomized setup. Therefore, this study will use an quasi-experiment setup.

There are four basic forms of a quasi-experiment (Millsap & Maydeu-Olivares, 2009), in this research we will use the 'One-Group, Pretest-Posttest Design'. The schematic visualization of observations and treatment are shown in Figure 9, where the 'O' indicates an observation and 'X' designates an intervention or treatment. In the case of this study the (learning) intervention 'X' will be the participation in Hazard Recognition as a controlled treatment condition.

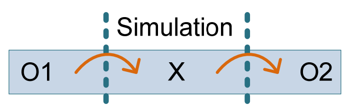

Figure 9 Observations and Treatment in 'One-Group, Pretest–Posttest Design'

Different than in a traditional One-Group setup is the fact that we can clearly distinguish two different kind of groups participating in the experiment. Therefore, an argument can be made that the groups are nonequivalent and should be treated as such. However, in a Nonequivalent-Group Design different participants receive a different (or no) treatment (Millsap & Maydeu-Olivares, 2009). In this study that is not the case, all participants receive the same treatment. The distinction between the two backgrounds should be considered as deviation in the moderating variables and how well the simulation performs for non-industry professionals.

3.4 Measuring Learning Efficacy

Measuring the learning efficacy is the main objective of this study. The learning efficacy is defined as the improvement of the participants' competences defined in the initial learning objectives of the serious game, in this case Hazard Recognition's learning objective is: *"To identify, assess and manage site hazards effectively"*. These competences are embedded in the pre – & post game conditions of the player and the learning efficacy of the game is defined by the change between both conditions. Thus, in order to operationalize Knowledge and Skills we need to measure the efficacy of the game in the achievement of its learning objectives in quantitative measurable factors.

The learning improvement of the game can be reported in two ways, the self-perceived competences of the players, or the player can be subjected to an objective test. In the first, the player reports on its competences from his/her own point of view, the latter test assesses actual factual knowledge and skills required for the supervising of hazards. The participant's report on their self-perceived competences is a participant's assumption and/or estimation of its set of competence and how they can be applied in reality. This means that a self-perceived measurement directly after playing the simulation would include the participant's assumption of the application of the newly acquired competences in future scenario's in real life. This is a valuable scale and very relevant in a study where participants are not easily subjected to new hazardous situations for the purpose of measurement. The self-perceived competences and the underlying theory will be elaborated more extensively in paragraph 3.4.1. On the contrary, an objective assessment would require an in-depth test and scenario in order to test the participant's skill improvement. Testing knowledge is somewhat easier, but still requires an extensive test. Regarding the scope and resources committed to this study it is impossible to assess participants extensively on their competences to assess and identify hazards. This would require extra real-life scenarios. In paragraph 3.4.2 there will be elaborated on the way to handle this issue.

As mentioned earlier, there is a distinction in the experimental group between participants with or without a background in the oil and gas industry. The expectation is that participants without a background in this industry fall in the *incompetence* area and participants with a background will fall in the *competence* area (Figure 10). The difference between *competence* and *incompetence* can be observed by the amount of knowledge and skill the participant has, measured on an objective scale. The distinction between *conscious* and *unconscious* can be discovered by the relationship between the scoring of the participant on the competence scale and the participants self-perceived competence.

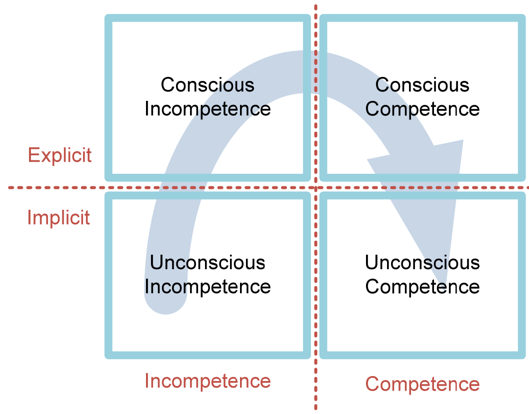

Figure 10 Competence Model (Noel Burch)

In line with the above logic the assumption is that when the self-perceived competence of the participant is measured low, but the competence is measured high, the participant is in the *unconscious competence* zone. All other combinations are shown in Table 3. The objective competence level of the participants is part of a larger competence classification used within Shell, there are four levels where a professional can operate in:

awareness → *knowledge* → *skill* → *mastery*

A combination of Shell's classification and the self-perceived competences would result in the conceptualization visualized in Figure 11. This model is a Shell specific derivative of the Competence Quadrant.

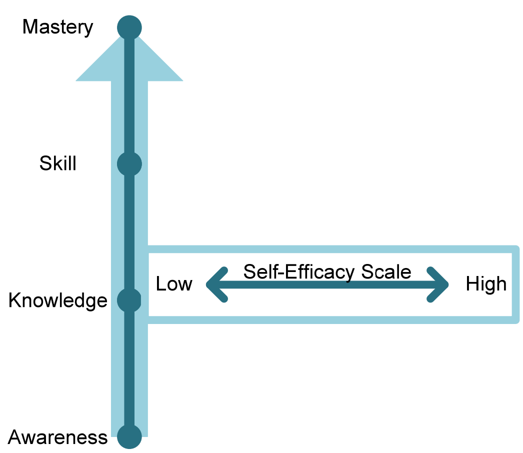

Figure 11 Self-Efficacy versus Mastery Level

Measuring the participants level on the scale classification used by Shell is not possible because of the scope and resources committed to this study. However, it can be derived on two indirect ways: the participant's objective performance in the workshop and assumptions based on the years of professional experience. This gives extra power to the objective assessment while placing the participants in the Competence Model Quadrant.

Competence Quadrant		
Competence	**Self-Perceived Competence**	**Quadrant**
High	High	Conscious Competence
High	Low	Unconscious Competence
Low	High	Unconscious Incompetence
Low	Low	Conscious Incompetence

Table 3 Competence Model Placement versus Score

3.4.1 Self Perceived Competences

Self-efficacy is the measure of one's own ability to complete tasks and reach goals (Ormrod, 2006) and affects every area of human endeavor. Self-efficacy is directly related to behavior. Perceived self-efficacy represents the confidence that one can employ skills necessary to mobilize one's resources required to meet the situational demand (Luszczynska & Schwarzer, 2005). Thus, a self-efficacy assessment on one's competences would be subjectively derived, however it is a useful scale to determine the perception a participant holds regarding its power to affect situations. Within Self-Determination Theory (SDT) self-perceived competence is assumed to be one of three fundamental psychological needs, so the feelings or perceptions of competence with respect to an activity or domain is theorized to be important both because it facilitates people's goal attainment and also provides them with a sense of need satisfaction from engaging in an activity at which they feel effective (Williams & Deci, 1996; Williams, Freedman, & Deci, 1998). According to Williams (1996; 1998) the Perceived Competence Scale (PCS) is a short, 4-item questionnaire, and is one of the most face valid of the instruments designed to assess constructs from SDT. Like several of the other measures, including the Self-Regulation Questionnaires and the Perceived Autonomy Support (Climate) Questionnaires, items on the PCS is typically written to be specific to the relevant behavior or domain being studied.

Reporting with self-perceived efficacy assumes a minimal level of the participant's ability to have self-awareness. This measurement approach can be translated into self-perceived competences, competences that relate to the learning objectives of the serious game. The perception of self-efficacy is a viable construct for comprehending performance. (Bouffard-Bouchard, 1990).

3.4.2 Objective Assessment

In addition to the self-efficacy assessment it is also important to gain insight in the level of knowledge and skill the participants gained measured on an objective scale. Such an observation contributes to the two dimensional area shown in Figure 11, which visualizes the two dimensional measurement. A participant could report a very high self-efficacy, but be actual low in real knowledge and skill. The importance of this observation and the application hereof has already been explained in the introduction of this chapter.

The objective measurements that are being conducted among the participants have two possible methods. The simulation has a build in scoring system, which tracks the players' achievements. These scores can be used to evaluate the players ability to tackle hazards within the simulation. However, this only shows how well a player performs, it obviously does not reveal the players improvement on this issue. That would require multiple comparable plays. Also, the scores are only logged by the computer when a player unlocks all achievements in the game and successively updates the scores manually. This might result in only a small amount of log files when participants are not able to finish a level or when they forget to store their score.

The second method is a verbal test on paper accompanying the measurement of self-perceived competences, this method allows to measure increase (or decrease) of objective knowledge on the subject. Because the group of participants has a distinct discrimination between professional backgrounds, the measurement tool should be identical for both groups for the sake of comparison, but should also be useful and meaningful for both groups. If the measurement tool does not contain meaningful context for the participant, results will be likely to be of poor quality. To achieve a meaningful context for both groups the questionnaire contains two types of questions:

a) Multiple choice questions about procedures and situations similar to situations in the simulation. This allows to measure if participants actually learned something directly from the game, without any knowledge beforehand, participants will receive information in the simulation on these topics. By asking the questions both before and after the game, a learning curve can be established and in general the simulation's stickiness of learning can be observed.
b) Recalling hazards from own experience or interpretation is the second method. The participant is asked to name up to 10 hazards that could be on a construction site. This also allows participants with a limited background in the industry to come up with their ideas about hazards or an interpretation thereof, while more experienced professionals

can name very specific hazards. This incorporates a flexibility towards all kind of participants whilst being relevant and meaningful for everyone. The amount of relevant entries are graded in order to achieve an objective scale. It will be taken into account that novices answer differently than professionals.

Thus, the combination of the three measurement tools covers an array of performances that try to balance the cognitive capacity of the player.

The in-game scores measure the ability to interact with the environment and the participant's awareness of hazards. The procedural questions test the player's knowledge base on relevant procedures, which are also present in the simulation. While the test on naming hazards does not only uses the knowledge base, but also requires to make abstractions and strategies because hazards can be so numerous it is a task of improvisation.

3.5 Conclusion on Experiment Design

This chapter was dedicated to a conceptual design of the experiment and to lay the foundations for the measurement tools that can establish knowledge on the learning efficacy of a serious game. Successively this will be used to measure Hazard Recognition's learning efficacy. The question that is proposed in this chapter, and which will be answered, is:

> *"How can the learning efficacy, and underlying factors, of a serious game be measured?"*

Because the learning efficacy of a serious game is defined by the movement of the student through the Competence Model due exposure to the learning intervention, there has been formulated a method to study this behavior. The experiment should be set up in a quasi-experimental fashion with two rounds of observations as described in the evaluation framework of Mayer et al. (2012). This framework defines the moderating variables that should be taken into account when investigating the underlying variables.

The learning efficacy, of the objectified competence(s), of the serious game is discovered by the use of a two dimensional measurement; the self-perceived competence (reflects the conscious axis) and the objective performance on the relevant competence(s) (reflects the competence axis). These measurements indicate the position of the student in the Competence Model (Burch, 1970s) and by measuring before and after the learning intervention, the movement can be exposed. For the assessment of the self-perceived competences, there can be made use of the Perceived Competence Scale: a 4-item questionnaire, one of the most face valid instruments within the domain of Self-Determination Theory (Williams & Deci, 1996; Williams et al., 1998). As discussed in the previous chapter, there should be sought at the most balanced assessment of the players' cognitive capacity, according to the Interactive Cognitive Complexion Model, to achieve the best overview of the players' learning achievements.

For the evaluation of Hazard Recognition the conceptual method has been made case specific. The experiment group has been selected and were divided into a clearly distinguished competent and incompetent group; professionals from Shell that were assumed to be relatively competent in assessing hazards, and players from Delft University of Technology and employees from Accenture, whom were assumed to be laymen in the field of hazard detection. With regard to the competences that Hazard Recognition aims to improve, Kowalski's model from the previous chapter should be used as foundation for both the self-perceived scale as for the objective performance. The players' objective performance will be measured with the use of in-game logging and pre- and post questionnaires; containing procedural and strategic questions.

4 Operationalization of the Method

In previous chapters the conceptual foundation has been formulated for the assessment of a serious game's learning efficacy. In this chapter the method will be further operationalized for the assessment of Hazard Recognition. Therefore, the research question that will be answered in this chapter is:

> *"How should the evaluation framework be operationalized for the evaluation of Hazard Recognition's learning efficacy?"*

4.1 Hypotheses

There have been formulated multiple hypotheses that offer a proposed explanation to the assumed outcomes of the experiment and contribute to the answering of this study's research question. These hypotheses will be assessed in this chapter by the use of statistical analysis on the data generated in the workshops. The data will be digested and analyzed with the use of Microsoft Excel and IBM's SPSS version 20. The main hypothesis for this research is:

> Hypothesis: *"After playing Hazard Recognition, players experience competence improvement for the detection and assessment of hazards"*.

In addition to the main hypothesis, there have been made several hypotheses that explain more detailed expected outcomes. These hypotheses and their explanation can be found in; Appendix IA, Appendix IB, Appendix IC, Appendix ID and are listed in Table 4.

Hypotheses

Subdivision		Hypothesis
	0A	*"There is a relationship between the player's self-perceived competences and the player's performance on an objective scale."*
Self-Perceived Competences	1A	*"Playing the simulation will cause a shift among all participants in their attitude towards hazards towards a more risk averse attitude."*
	1B	*"There is a positive relation between the attitude towards hazards measured before playing the game and the participants self-perceived competences measured both before and after the game."*
	1C	*"There is no difference in the increase in self-perceived competence between participants with or without a background in the oil & gas industry."*
	1D	*"Participants with a background in the oil & gas industry have an inverse relationship between years of experience in the industry and their level of self-perceived competences (both before as after the simulation)."*
	1E	*"Individual factors such as sex, nationality and highest level of education have no effect on the level of self-perceived competences and the amount of increase of the self-perceived competences."*
	1F	*"There is a positive relation between the level of immersiveness of the participant and the increase of their self-perceived competences."*
Objective Performance	2A	*"There is a positive relation between the attitude towards hazard measured before playing the game and the participants objective performance measured both before and after the game."*
	2B	*"Participants with a background in the oil & gas industry have a better objective performance than participants lacking this background."*
	2C	*"There is a positive relation between the amount of years a participant has worked in the oil & gas industry and their objective performance."*
	2D	*"Individual factors such as sex, nationality and highest level of education have no effect on the objective performance nor on the objective learning efficacy."*
	2E	*"Participants that experience a higher degree of enjoinment of learning about hazards have a higher objective performance than participants who enjoined this less."*
	2F	*"There is a positive relation between the level of immersiveness of the participant and the objective performance."*
Immersiveness	3A	*"Participants who have a higher regard of the simulation's design enjoy a higher degree of immersiveness."*
	3B	*"Participants without a gaming background will enjoy an average feeling of immersiveness, participants with a gaming background will enjoy either an above average immersiveness due to their experience in the use of games or an beneath average due to their critical viewing of the simulation."*
	3C	*"The interference of structural bugs will have a negative impact on the participants level of enjoyment playing the game."*
	3D	*"There is a positive relationship between the participant's level of learning fun and the level of immersiveness the participant enjoys."*
	3E	*"Participants who have a good perception of the balance between the in-game reward structure and the difficulty of achieving this, enjoy a higher immersiveness than participants with an off-balance perception."*

Table 4 Hypotheses

The logic behind the analysis and these hypotheses is that there can be looked to differences in results based on the background of the participants, the relation between learning efficacy outcomes and participants' score on assumed explaining factors, and a combination of the two. Now this is clear we know what the learning efficacy is of Hazard Recognition in general, in which extent this efficacy is achieved among the different backgrounds and which factors contribute hereto. The assumed relationships between the different concepts are shown in; Figure 12, Figure 13 and Figure 14. Where the red boxes are the two defining indicators on the learning efficacy of Hazard Recognition. These relationships are derived from the Comprehensive Framework and literature (Alvarez, 2008). The factor *immersiveness* is dark blue to indicate its bridge function between other underlying factors and the learning efficacy indicator.

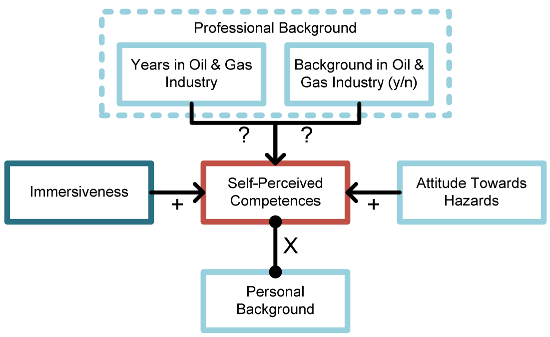

Figure 12 Hypotheses Relationships Self-Perceived Competences

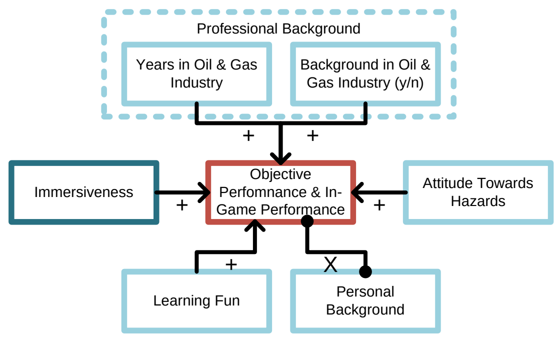

Figure 13 Hypotheses Relationships Objective Performance

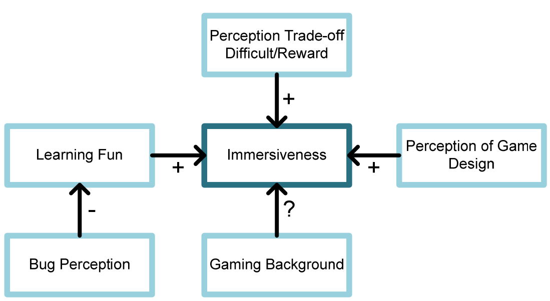

Figure 14 Hypotheses Relationships Immersiveness

4.2 Method Construction: Tools and Workshop

In previous paragraphs it as accurately proven that the logic behind the factors we want to measure in order to establish knowledge that can fund answers to the main research question. In this paragraph there will build on the knowledge derived earlier in this chapter, resulting in measurement tools and an experiment setup. The constructs we are interested in can be distinguished in four categories: Self-Perceived Competences, Objective Performance, Experience and Background Data. With the operationalized framework in mind and the competences we want to measure, the items that will be measured are shown in Table 5.

Measurement Items			
Perceived Competences	*Objective Performance*	*Background Data*	*Game Experience*
5 Competences Kowalski	In-Game Performance	Age, Sex, Nationality	Immersion
Perceived Knowledge	10 Hazards on Worksites	Level of Education	Learning Fun
	Procedural Questions	Education Disciplinary	Trade-off Difficult/reward
		Experience in Industry	Perceived Game Design
		Gaming Background	Bug Perception
		Attitude towards Every Day Hazards	
		Professional Hazard Experience	

Table 5 Measurement Items

There are several ways to measure the above-mentioned items. In this study, time is an important constraint, because the simulation alone takes at least an hour. So, in order to present a workshop that is considerable to the participants and their hosts' schedule, the measurement tools should not take the participants more than 30 minutes of their time. A workshop could be finished within 90 – 120 minutes.

The operationalized framework has a clear view on when which measurements should be conducted. Considered the resources committed to this study, data can be gathered by: in-game log files, questionnaire and video. This means that all items from Table 5 should be matched with at least one of the three measurement tools. Figure 15 shows the conceptual model for the appropriate place of these observation methods according to the experiment design (chapter 3.3).

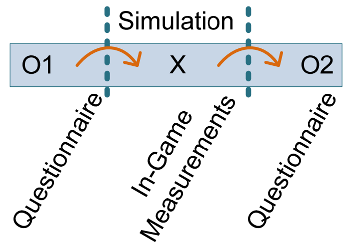

Figure 15 Operationalization of 'One-Group, Pretest–Posttest Design'

When looking at the nature of the measurement tools, a distinction can be made between questionnaire on one side and video and log files on the other. The questionnaire is a method that requires the participant's attention, while the others are passive registration methods that could be used recording game play. Let us assign the items accordingly to the appropriate methods.

- First, the in-game log files can track the player's failure, i.e. when a hazard fires this will be logged. Thus, the in-game performance can be logged with these files.
- Second, videotaping participants can most predominantly assess two things:
  - The behavior of the whole classroom and their engagement,
  - Or when zoomed in on one or two players it can track very closely their immersiveness. The limitations however, are the amount of camera's. So, the measurements for these factors should not rely on camera alone.
- Third, the questionnaire is a technically simple, but powerful tool to collect information. Before the constructs of the items are elaborated more extensively, in Table 6 an extend overview of the method and placement of the measurements in the workshop, in line with the operationalized framework, is shown.

Measurement Items		
Before the Simulation	*During the Simulation*	*After the Simulation*
Age, Sex, Nationality	In-Game Performance (log files)	5 Competences Kowalski
Level of Education	Immersion (video)	Perceived Knowledge
Education Disciplinary		Attitude towards Every Day Hazards
Gaming Background		10 Hazards on Worksites
Experience in Industry		Procedural Questions
Professional Hazard Experience		Immersion
5 Competences Kowalski		Learning Fun
Perceived Knowledge		Trade-off Difficult/reward
Attitude towards Every Day Hazards		Perceived Game Design
10 Hazards on Worksites		Bug Perception
Procedural Questions		
Questionnaire	**Log Files & Video**	**Questionnaire**

Table 6 Placement of Measurements in Workshop

The constructs that build up the total questionnaire can be derived from existing validated questionnaires, which saves time and effort, or are to be constructed. A list of already validated scales that will be used in the questionnaire:

a) The perceived competence scale and its origin has been elaborated on in paragraph 3.4.1 and will be applied to the competences from the Kowalski Model (2003) and on the perceived knowledge of the participant.
b) An immersiveness scale has been developed and validated by (Brockmyer et al., 2009).
c) A validated scale for the perceived quality of the game design is available (TU Delft Knowledge base)
d) A set of questions that define the gaming background of the participant derived from the TU Delft knowledge base.

The other constructs have been designed as part of this study, after the data has been gathered the newly developed scales will be tested on their internal reliability. It is not necessarily so that the newly developed constructs derive the correct data of the subject of interest. Hence, if the new constructs does or does not measure a relation or distinction, it does not directly imply the validity of the construct. With this in mind, the new constructs are developed with simplicity and objectivity in mind, in addition questions are not implicitly asked. With objectivity of constructs is meant that factual questions are asked and not interpretations or opinions. Simplicity means that questions are straight forward and do not appear as trick questions. Both values are hopefully strong enough to derive strong and reliable data. Table 7 shows the new developed constructs.

Except for the objective performance constructs, all of them are measured on a 1 to 5 Likert scale.

Developed Constructs		
Objective Performance		
Hazard Awareness	Q1	Name Maximum 10 types of typical hazards on worksites:
Procedural Knowledge	Q1	The banksman should have eye contact with? a) Crane driver, b) lifting personal, c) supervisor, d) no one
	Q2	Who is responsible for signaling the cranedriver? a) Banksman, b) Supervisor, c) Lifting Personall, d) no one
	Q3	With heavy winds, you can work on a: a) Cherry Picker, b) Ladder, c) Crane, d) Not on Heights
	Q4	Lifting straps should be renewed every … a) 6 months, b) 12 months, c) 18 months, d) 24 months
	Q5	The banksman is allowed to … a) Giving hoisting directions, b) Smoke, c) Help with the load, d) Ask for permits
Background Data		
Professional Hazard Experience	Q1	In my daily work, I'm actively involved in hazardous situations?
	Q2	In my daily work, I'm actively involved in hazard detection?
	Q3	In my daily work, I'm actively involved in responding to hazards?
	Q4	Before I took part in this simulation game, I was already experienced and informed about issues related to hazards
	Q5	I'm a subject expert in my field of profession
	Q6	Hazards are part of my every day working environment
Attitude towards Hazards	Q1	I would always secure a ladder when using it
	Q2	I would use fall protection for cutting high trees
	Q3	I would ask someone to be an extra pair of eyes for cutting complex tree structures
	Q4	I would use a helmet when hoisting a piano to the second floor's window
	Q5	I would first thoroghly inspect hoisting equipment before using it
	Q6	I would use an improvised construction of a table and chair to change a light bulb
Game Experience		
Learning Fun	Q1	The game was fun
	Q2	The exposure to a different industry (chemical sites) was interesting
	Q3	It was fun to learn things about hazard recognition
	Q4	I learned much about the recognition of hazards
Trade-off Difficulty/reward	Q1	The tasks in the simulation game were too easy.
	Q2	The tasks in the simulation game were too difficult.
	Q3	The structure of the game was too predefined.
	Q4	I experienced a balance between difficulty of the tasks and rewards I gained
Bug Perception	Q1	In my perception, the game was bug free
	Q2	The bugs I noticed were of negative influence on my performance as Supervisor
	Q3	The bugs I noticed took me out my game flow
	Q4	Yes I noticed bugs, but they didn't influenced my gameplay or fun

Table 7 Developed Constructs

In Table 8 is shown an overview of the used constructs, their names as used in the analyses, their meaning and the internal reliability as they were tested in the experiment. The experiment will be elaborated on in the next chapter, but for the sake of perspicuity the internal reliability and number of cases is already presented here.

Constructs				
Name	*Meaning*	*N*	*Crombach's Alpha*	
Background Gaming	The value one gives to the use of serious games in education	31	0,719	
Professional Background	Perception of what degree the professional environment is shaped by hazardous situations	56	0,697	
Attitude_(Pre/Post)*	Objective measurement on the risk averseness in private capacity	85|88	0,785	0,838
Detect_(Pre/Post)*	Self-perceived competence of detecting hazards (derived from Kowalski's model)	82|87	0,832	0,949
Recognize_(Pre/Post)*	Self-perceived competence of recognizing hazards (derived from Kowalski's model)	85|87	0,869	0,949
Respond _(Pre/Post)*	Self-perceived competence of responding on hazardous situations (derived from Kowalski's model)	85|88	0,908	0,963
Control_(Pre/Post)*	Self-perceived competence of controlling hazardous situations (derived from Kowalski's model)	54|57	0,913	0,957
Oversight_(Pre/Post)*	Self-perceived competence of keeping oversight on the hazardous situation (derived from Kowalski's model)	54|54	0,943	0,932
Explain_(Pre/Post)*	Confidence of a participant to explain on hazards on worksites, i.e. sharing knowledge	55|85	0,855	0,911
Procedural_(Pre/Post)*	Objective assessment on procedural knowledge, procedures that are in Hazard Recognition	48|76	-	
10 Hazards_(Pre/Post)*	Open question, required to list up to 10 hazards that could be present on a worksite	56|80	-	
Game Design	Participants valuation of the quality of the game design	31	0,902	
Immersion	Level of immersiveness the participant experienced	29	0,785	
Trade Off	Player's valuation of the reward versus difficulty balance of the game	30	0,330	
Learning Fun	Only asked to non-industry participants, what the level of enjoyment was to learn about a different industry and the amount of fun the game was	30	0,793	
Bug Perception	The perception of gameplay interference due to bugs in the game	11	0,763	

***(Pre/Post) means that the construct is measured before and after the game → The difference is noted as '<construct>_Delta'**

Table 8 Constructs and their Meaning

4.3 Conclusion Operationalization of the Method

This chapter focused on the construction of a measurement method, derived from the previously constructed conceptual model. The question that shaped this chapter's process is:

> *"How should the evaluation framework be operationalized for the evaluation of Hazard Recognition's learning efficacy?"*

By the means of formulating and testing multiple hypotheses the learning effects of Hazard Recognition and underlying contributing factors will be studied. The hypotheses focus on three areas; Self-Perceived Competences (Figure 12), Objective Performance (Figure 13) and Immersiveness (Figure 14). Immersiveness is taken into account specifically because literature indicates its bridge function between performance and underlying game(play) factors (Alvarez, 2008).

The measurement items that will establish knowledge about the hypotheses can be divided in four categories; Perceived Competences, Objective Performance, Background Data and Game Experience (Table 5). The items will be measured in a quasi-experimental fashion by the use of questionnaires and in-game measurements (video logging and in-game scoring logs), as ordered in Table 6.

The constructs that will be used in the questionnaires are derived from literature (Williams, 1996; 1996; Brockmeyer et al., 2009; Kowalski, 2003), gathered from the TU Delft's knowledge base or are newly constructed (Table 7). The result is the questionnaire found in Appendix V, an overview of the constructs is found in Table 8.

5 Experiment Conduct

After the operationalization of the evaluation framework and the construction of the measurement tools the data gathering can commence. The data about the learning efficacy and the explaining factors will be gathered by the use of workshops capacitated for 15 to 48 participants. There have been organized 4 workshops at Royal Dutch Shell NV., Shell Aviation Canada, Delft University of Technology and Accenture Netherlands, this amounted to the total of 91 participants. 60 Of which with a professional background in the oil & gas industry. The total participant group consisted of around 15% females and around 85% males, these people held 22 different nationalities. The educational background differs from high school graduates (~8%) up to PhDs (~4%), the majority has a MSc degree (52%). This very diverse group of people have been subjected to an experiment aiming to answer the main research question and in particular this chapter's question:

"Does Hazard Recognition achieves learning efficacy and what factors contribute hereto?"

5.1 Experiment Outcome

The hypotheses have been tested and will be reported upon in this sub-chapter. In this research a significance level of 95% has been used, which is a standard for this type of research. Because of the chronology of this research the measurement methods have been subjected to iteration and evolved on different accounts.

The data that has been derived during the workshops have been digested in order to use it for further analysis. This digestion consists most commonly of recoding variables into their inverse properties so that all variables in the same construct point in the same direction. Also, variables have been merged when necessary. Before merging variables, the individual parameters have been tested on their internal reliability. All scales (also the merged scales) are from 1 to 5, in the explanation 1 means poor and 5 means good. For the hypotheses where a relation between constructs is proposed there has been made use of a curve fitting method using five possible curves: linear, quadratic, cubic, s-curve, growth. The reason behind this method is that there is no indication that all relations are per se linear, the accumulation of knowledge is for instance something that might smooth out over the years. This method offers a broader scope on possible relations and suits the purpose of this research.

5.1.1 Self-Perceived Competences and Objective Performance

Hypothesis 0A: *"There is a relationship between the player's self-perceived competences and the player's performance on an objective scale."*

Because both the self-perceived competences as well as the objective performance is measured in a multitude of constructs, the exploration will focus on the relationships between the independent scales rather than on the sum of both. This gives a clear understanding of the independent relations and neglects the damping of effects that would be observed on a higher aggregation level (the two competences). The results are shown in Table 9 and Table 10. What is interesting about this observation is that of the three construct defining the objective performance, only the construct defining the ability to name hazards correlates with constructs of the self-perceived performance. In addition, there is no relationship witnessed among non-oil & gas professionals, but a combination of both groups results in more significant relationships than observed in the Shell group alone.

Self-Perceived Competences versus Objective Performance (All participants)				
Self-Perceived Competences	*n*	*Open Question Pre*	*Open Question Post*	*Open Question Delta*
Recognition_Pre	77	-	0,284 [0,081] (0,012)	-
Responding_Pre	77	-	0,334 [0,111] (0,003)	-
Detect_Post	78	-	0,342 [0,117] (0,002)	-
Recognition_Post	78	-	0,339 [0,115] (0,002)	-
Responding_Post	79	-	0,236 [0,055] (0,037)	-
Explain_Post	79	-	0,388 [0,150] (0,000)	-
Pearson's Correlation [R²] (significance level, 2 tailed)				

Table 9 Competence Correlation Matrix All Participants

Looking at the correlations between the constructs, what comes directly to mind is the relative low values of R^2, which indicate a relative low amount of predictive power between the constructs and a low goodness of fit. A correlation measures per definition a linear relationship, this would leave room for a relationship of a different curve with a higher R^2. However, the scatterplots of the relationships show no indication that a curve would fit better (Appendix IIA). The distribution of the cases is just really wide and prove no real relationship.

Regarding the significant correlations measured in the sole Shell professionals group, we can conclude that also in this group the R^2 is relatively low, with the same consequences as described above.

Self-Perceived Competences versus Objective Performance (Shell Participants)				
Self-Perceived Competences	*n*	*Open Question Pre*	*Open Question After*	*Open Question Delta*
Responding_Pre	51	-	0,418 [0,175] (0,002)	-
Detect_Delta	52/50	-0,353 [0,125] (0,010)	-	0,395 [0,159] (0,005)
Responding_Delta	53	-0,378 [0,143] (0,005)	-	-
Pearson's Correlation [R²] (significance level, 2 tailed)				

Table 10 Competence Correlation Matrix Shell Participants

Thus, we can conclude that a selection of the self-perceived competences show relationships with the ability of participants to name hazards, which reflects awareness and knowledge of hazards. In addition, there seems to be no relation between self-perceived competences and the in-game performance, nor on the multiple choice questions about hazard prevention procedures.

5.1.2 Self-Perceived Competences

Hypothesis 1A: *"Playing the simulation will cause a shift among all participants in their attitude towards hazards towards a more risk averse attitude."*

With an significance of 99% it can be said that the attitude of players will shift towards more risk averse attitude after playing the game. Shifting the mean from 3,65 up to 3,87.

If we differentiate between participants with or without a background in the oil and gas industry, we find that both groups increase their risk averseness in the same absolute amount, but that the level of a risk averse attitude is higher with participants with a background in the oil and gas industry (see Table 11 for results).

Results Hypothesis 1A			
Group	*Construct*	*n*	*Mean*
All Participants	Attitude_Pre	84	3,65
	Attitude_Post	84	3,87
	Attitude_Delta	84	0,22
Oil & Gas Background*	Attitude_Pre	54	3,85
	Attitude_Post	57	4,07
	Attitude_Delta	53	0,21
No Oil & Gas Background*	Attitude_Pre	31	3,28
	Attitude_Post	31	3,52
	Attitude_Delta	31	0,24
***Difference between groups on a 99% confidence scale (p < 0,001)**			

Table 11 Results Hypothesis 1A

Hypothesis 1B: *"There is a positive relation between the attitude towards hazards measured before playing the game and the participants self-perceived competences measured both before and after the game."*

Only the linear correlation between Attitude and 2 self-perceived competence Recognition measured after the simulation and Responding (before and after the simulation) is significant (Table 12). However, the R^2 is so low (all under 0,085) that a relationship cannot be assumed.

Self-Perceived Competences versus Initial Attitude towards Hazards		
Self-Perceived Competences	*n*	*Open Question Pre*
Recognition Post	82	0,293 [0,068] (0,008)
Responding Pre	84	0,267 [0,063] (0,014)
Responding Post	83	0,316 [0,084] (0,004)
Pearson's Correlation [R^2] (significance level, 2 tailed)		

Table 12 Self-Perceived Competences versus Initial Attitude towards Hazards

Hypothesis 1C: *"There is no difference in the increase in self-perceived competence between participants with or without a background in the oil & gas industry."*

The three self-perceived competences that have been measured in the whole experiment group: Detection, Recognition, Responding, show an interesting trend. Only in the non-industry group on Detection experiences a decrease in self-perceived competence, everything else went up in both groups. Not surprisingly, for Detection there has been observed a significant difference in the delta (difference between before and after playing the simulation) between both groups. On the competence Recognition the significance level was only 92%, therefore a significant difference may not be assumed. However, the absolute difference between both groups is big; 0,28 (industry group) versus 0,00 (non-industry group), considering the initial values of the competence. There has to be taken into account that the original scale is from 1 to 5, i.e. an increase of 1,00 is a 20% increase of the whole scale.

Thus, the increase in self-perceived competences do or have the tendency to differ between industry groups, but the measurements do not support this with full confidence. What these results also show, is that novices have a diminishment in their confidence of the detection and stagnation in the recognition of hazards. But their self-perceived competence of responding to hazards grows. This means that novices feel that once a hazard is found, they would be able to handle it with more confidence after playing Hazard Recognition.

Results Hypothesis 1C				
Construct	*Oil & Gas Industry*	*n*	*Mean*	*Difference Significant**
Detect_Delta	Yes	52	0,23	0,005
	No	28	-0,29	
Recognition_Delta	Yes	53	0,28	0,098
	No	30	0,01	
Respond_Delta	Yes	53	0,25	0,576
	No	31	0,15	

*** significance level, 2 tailed, difference measured between groups**

Table 13 Results Hypothesis 1C

Hypothesis 1D: *"Participants with a background in the oil & gas industry have an inverse relationship between years of experience in the industry and their level of self-perceived competences (both before as after the simulation)."*

There is no overwhelming evidence that the level of self-perceived competences is in any case dependent on the years of experience a player has in the oil & gas industry, not on individual competence level as on a combined scale.

Participants with an oil & gas background have been tested on five self-perceived competences. There could not be observed a convincing relation between years of experience and the combination of the five self-perceived competences before playing the game. A linear significant regression was measured for the combined scale of competences measured after playing the game, but the goodness of fit (R-square) was only 0,093. Therefore the measured relationship is negligible.

Not really different is the relation for the self-perceived competences observed after the simulation. A nearly significant cubic regression ($p = 0{,}051$) was measured, displaying a slightly higher goodness of fit (R-square = 0,161).

In addition all five competences have been tested individually. The results were also not suggesting a strong relationship between years of experience and the perception of competence. Only the competence Responding (before simulation) was tested significant on a linear regression, however the goodness of fit was lower than 0,09. This is unacceptable low figure. The second significant ($p=0{,}043$) relationship observed is that with Control (after simulation) on a cubic curve with a goodness of fit of 0,16. Also a low explained variance. The reason that a nonlinear regression method is used, is that a saturation in competence might occur over the years. This would indicate nonlinear behavior. Another possibility is that at some point in time professionals have seen so much hazards and deviations in dangerous situation in the industry, that they become more modest in their valuation of competences.

Hypothesis 1E: *"Individual factors such as sex, nationality and highest level of education have no effect on the level of self-perceived competences and the amount of increase of the self-perceived competences."*

The level self-perceived core competences (Detection, Recognition and Responding) are distributed evenly between both sexes, when looking to the increase of the competences only a significant difference can be observed with the competence Detection. Females show a decline in their self-reported competence Detection of -0,31 where man increase on average with 0,10. However, there has to be taken into account that due to the industry characteristics only data of 11 females has been collected. This means that a strong conclusion cannot be drawn.

The education level of the participants ranged between high school degree up to PhD, with also 2 participants with a professional aviation education. However, the weight of the distribution lied in the categories Bachelor and Masters (combined 80%). For this reason College and High School degrees would be added to the Bachelor population and the four PhD's would be added to the MSc group. This resulted in two balanced groups that can be compared. The analysis shows that the BSc group perceives their self-perceived competences on average higher than the MSc group. The increase in these levels is higher within the MSc group. However, only the self-perceived competence Recognition is significant (Table 14).

Education Level

Self-Perceived Competence	*Edu*	*N*	*Mean*	*Difference Significant**
Detect Pre	MSc	42	3,75	0,138
	BSc	39	3,92	
Recognition Pre	MSc	45	3,49	0,022
	BSc	39	3,78	
Respond Pre	MSc	45	3,65	0,130
	BSc	39	3,83	
Detect Post	MSc	44	3,73	0,091
	BSc	40	4,01	
Recognition Post	MSc	43	3,70	0,192
	BSc	41	3,92	
Respond Post	MSc	44	3,89	0,266
	BSc	41	4,06	

* significance level, 2 tailed, difference measured between groups

Table 14 Self-Perceived Competence divided by Education Level

Although the total population consisted of participants holding a variety of 22 nationalities, the greater majority was Dutch (33%). The rest of the participants were more or less evenly distributed amongst the 21 remaining nationalities, resulting in very small subgroups which left no room for statistical analysis.

> Hypothesis 1F: *"There is a positive relation between the level of immersiveness of the participant and the increase of their self-perceived competences."*

None of the three core-competences show any relation between the improvement thereof and the level of immersiveness of the participant.

Conclusion Self-Perceived Competences

The findings of this paragraph are that:

a) Playing Hazard Recognition creates a more risk averse attitude among participants;
b) Participants with a BSc degree value their competences higher than participants with a MSc degree. This could be due to the nature of working more on site, the underlying reason is unknown.
c) Years of experience in the oil and gas industry has no influence on the perception of ones competences.
d) Novices experience a diminishment in their confidence of the detection and stagnation in the recognition of hazards. But their self-perceived competence of responding to hazards grows. This means that novices feel that once a hazard is found, they would be able to handle it with more confidence after playing Hazard Recognition. Thus, the process steps before the intervention on a hazard are in a valley of despair, while the assessment thereof the opposite.

5.1.3 Objective performance

The objective achievements of the players can be decomposed in three items: 1) The in-game achievements, 2) The ability to sum up hazards, 3) The identification of procedural pinpoints in a multiple choice test.

There are two ways of looking at those variables in the search for relationships between the objective performance and other factors; looking independently at each factor or combine all factors together into one scale. If there would be a good way of combining the variables into one scale, that would make the explanation of relationships easier. However, all three items measure something else. Whereas the summing up of hazards more relates to the awareness and present knowledge on situations, measure the questions about procedures more the present knowledge. Also, the in-game achievements tell not only a story about the player's skills, but also say something about the ability of the player to cope with the simulation's environment and interaction.

The in-game performance of the participants is measured according to the amount of hazards that fire in each level. This means that a low score equals a good performance and vice versa. According to the measurements the Lifting and Hoisting level was more difficult than the Working on Heights level, there is a three time higher chance that a participant will fail in detecting and managing a hazard in the first level compared to the latter (μ=0,28 versus μ=0,92 per hazard).

> Hypothesis 2A: *"There is a positive relation between the attitude towards hazards measured before playing the game and the participants objective performance measured both before and after the game."*

If we look at the perceived relationship between the independent factor attitude towards hazards on commencement of the simulation and the ability of players to name hazards both before and after their participation, we can distinguish a positive relationship with a goodness of fit of respectively 12% and 18%. However, there is a significant difference between the levels of attitude between oil and gas professionals and laymen (Table 11). When we exclude the laymen in the analysis the goodness of fit declines to 6%. When only including laymen in the analysis this value is 4%. Which are both negligible values. Thus, the total experiment group shows some relevant regression, when broken down between industry expertise the effect is gone.

Attitude versus Objective Performance				
Attitude versus	*n*	R^2	*Regression Coefficient*	*Significance*
10 Hazards Pre	53	0,118	0,344	0,11
10 Hazards Post	77	0,174	0,417	0,000

Table 15 Results Hypothesis 2A - Regression Attitude versus Naming Hazards

The initial attitude towards hazards has no relationship with the performance of participants on the procedural questions both before and after their participation.

The in-game performance which is measured in both levels shows no significant relationship with the initial attitude towards hazards of the players. The relationship between failure and attitude in the Working on Heights level is an insignificant flat line. For the more difficult level Lifting and Hoisting the scatterplot shows a relative steep upward direction, with a proposed explanation of variance of 13%. But, this relationship is not significant. What is remarkable is that when this relationship would have been significant, that would have meant that a more risk averse initial attitude would be correlated with the relative inability to detect and manage hazards.

Hypothesis 2B: *"Participants with a background in the oil & gas industry have a better objective performance than participants lacking this background."*

Of the six variables that have been measured overall, only two have been measured in both groups. So a comparison between industry professionals and laymen has been made on the participant's performance on the procedural multiple choice questions and the ability to sum up worksite hazards, both after the simulation. No difference has been observed between the groups on their performance answering the procedural questions (both groups had an average score of 0,76). However, a significant difference in performance is observed in the summing up of worksite hazards after playing the simulation. Industry professionals are able to sum up on average 8,7 hazards, whereas laymen are able to name 6,1 on average.

Hypothesis 2C: *"There is a positive relation between the amount of years a participant has worked in the oil & gas industry and their objective performance."*

There has been no relationship observed between any performance indicator and the years of experience in the oil & gas industry of the participants. The scatterplots show a widespread deviation, which is evenly distributed (Appendix IIB). This indicates that the assessments have no distinctive power regarding the years of professional experience.

Hypothesis 2D: *"Individual factors such as sex, nationality and highest level of education have no effect on the objective performance nor on the objective learning efficacy."*

For the testing of this hypothesis, the same combination strategy has been used as has been done with hypothesis 1E. Following the same rational. The use of this strategy resulted in the observation that there are no differences between both education groups in the objective performance, except for the summing up of hazards on worksites after playing the simulation. A significant ($p<1\%$) difference was observed, the MSc group was able to sum up on average 7,1 hazards. Whereas the BSc group was able to sum up 8,6 hazards on average. This difference can have two main reasons:

First, the non-industry group consisted of almost only people with a MSc title. This could result in an unfair off-set in the results.

Second, it could be possible that a typical industry job with a BSc background is more involved with onsite hazards than a MSc job. However, with the data available only the first concern can be checked. By only taking into account the cases with an oil & gas background there is no significant difference observed in any objective performance between BSc and MSc education level.

As discussed in the previous sub-chapter, there is a widespread set of nationalities with very small cases, which leaves no room for a statistical analysis. Therefore there cannot be drawn any conclusions on the objective performance and the nationality of the participants. Also, there has not been observed a significant deviation in objective performance between the sexes. The female participant base is very small, which contributes to the inability to compare both groups.

These observations are of the performance levels before and after the simulation. They say nothing about the perceived improvements, i.e. efficacy of the game. When looking at the performance increase (i.e. the delta of the constructs) of the participants on the identification of hazards and the procedural questions, there is no distinction observed between the sexes nor between the education levels.

When looking at the distribution of the improvement on the performance indicators (see Figure 16 and Figure 17), we see that both indicators have the majority of the distribution weight on 0,0. Meaning that there is little improvement observed. These improvements have only been measured in the oil and gas professional group.

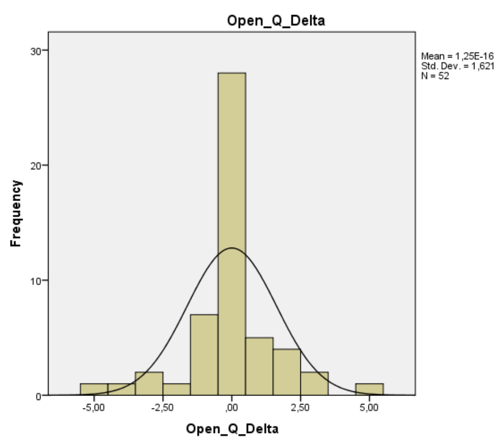

Figure 16 Frequency Chart – Increase Performance Naming Hazards (Shell Participants)

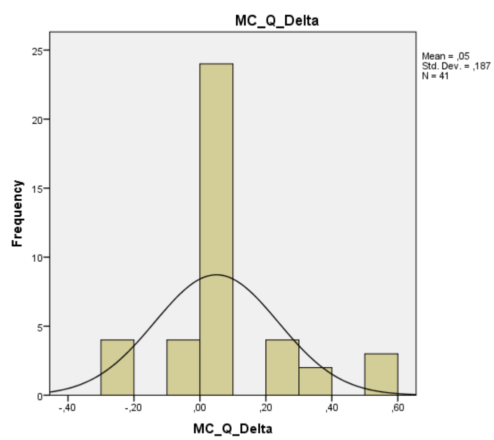

Figure 17 Frequency Chart – Increase Performance Procedural Questions (Shell Participants)

Hypothesis 2E: *"Participants that experience a higher degree of enjoinment of learning about hazards have a higher objective performance than participants who enjoyed this less."*

The amount of fun participants enjoyed during the learning about hazards was only measured amongst the laymen group and is close to normally distributed, with a mean around 3,6 (out of 5). This means that the average novice participant enjoyed the learning aspect of the workshop and enjoyed learning about hazards in an industry they are unfamiliar with. However, there is no relationship observed between the amount of learning fun the participants experienced and their objective performance. Thus, participants who experience little learning fun, are likely to perform on the same level as participants with a high degree of learning fun.

> Hypothesis 2F: *"There is a positive relation between the level of immersiveness of the participant and the objective performance."*

The level of immersion of the participants has no influence on their level of objective performance, measured on the two indicators Hazard Identification and the Procedural Questions.

Conclusion Objective Performance

This paragraph's findings are as follows:

a) The initial attitude towards hazards of participants has a positive relationship with the ability to sum up hazards, which reflect awareness. Not surprisingly, professionals score better with an average of 8,7 hazards, where laymen are able to sum up 6,1 hazards on average.
b) The initial attitude towards hazards is not a predictor for the ability to perform good on procedural assessments, nor on the in-game hazard assessment performance of participants.
c) After playing the simulation, novices score as good as professionals on the procedural assessment. This would indicate that the simulation is a powerful method to teach novices hazard procedures.
d) The overall objective performances of industry professionals does not change depending on the amount of years the participant has worked in the oil and gas industry.
e) Within the oil and gas professional group, little increase is observed in their objective performance.
f) The average non industry participant enjoyed the learning aspect of the workshop and enjoyed learning about hazards in an industry they are unfamiliar with (average of 3,6 on a 5 point scale). This indicates that the simulation is an enjoyable method for novices in learning industry specific subjects.

5.1.4 Immersiveness

Hypothesis 3A: *"Participants who have a higher regard of the simulation's design enjoy a higher degree of immersiveness."*

There seems to be a correlation between the level of immersiveness the player experiences and the player's regard of the simulation its design quality (scatterplot in Figure 18). Because the causality between the two variables is unclear, we can only conclude that there is a relation (Pearson's coefficient 0,531, explained variance of 28%, p = 0,002). This causality is unclear because it is plausible that when someone is more immersed this player regards the game design quality higher, but the other way around is also possible; someone who finds the game design poor is unlikely to be immersed.

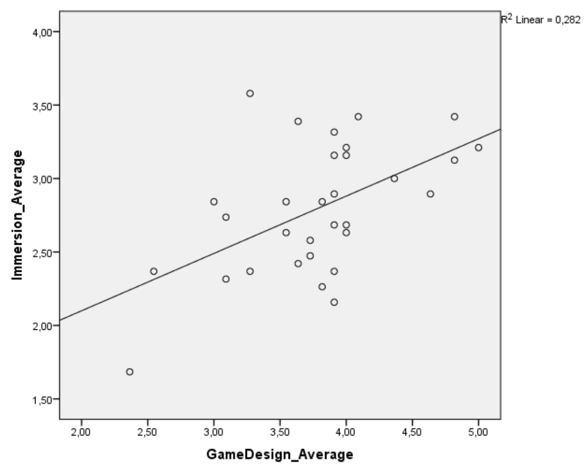

Figure 18 Correlation Game Design Valuation versus Player's Immersiveness

Hypothesis 3B: *"Participants without a gaming background will enjoy an average feeling of immersiveness, participants with a gaming background will enjoy either an above average immersiveness due to their experience in the use of games or a beneath average due to their critical viewing of the simulation."*

The immersiveness is distributed relatively wide among the participants. When looking at the distinction between the gaming and non-gaming background groups we can observe a similar distribution in both groups (see Figure 19). Both groups have similar means (both means around 2,8), only a little higher standard deviation for the gaming background group (0,47 versus 0,39), however according to Levene's Test equal variance should be assumed. This means that the distinction presumed in this hypothesis is false.

Also, the difference in experience of participants with business games showed no distinction between the groups and the level in which the player is immersed in the simulation.

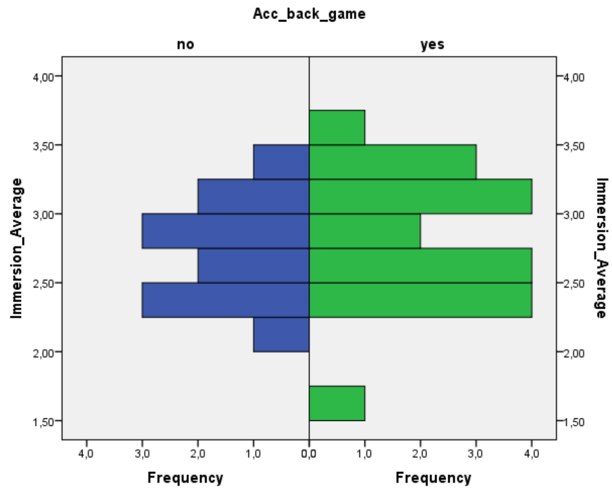

Figure 19 Hypothesis 3B - Distribution of Immersiveness between Gaming Background (y/n)

Hypothesis 3C: *"The interference of structural bugs will have a negative impact on the participants level of enjoyment playing the game."*

The analyses on this subject showed that there is a relationship between the level of perceived structural consistency in the game and the level of enjoyment in learning the topic of hazard recognition. Participants who perceived more bugs had a lower level of learning fun. However, due to workshop constraints the amount of cases for this analysis was limited to n=11, which makes this result statistically unreliable.

In addition, the same parameter on bug perception has been investigated on its relationship with the participants immersiveness, this analysis showed no relationship.

Hypothesis 3D: *"There is a positive relationship between the participant's level of learning fun and the level of immersiveness the participant enjoys."*

There is a strong relationship (Table 16 and Figure 20) between the amount of fun the participant experiences on learning about the subject of hazard recognition and the amount of immersiveness the player enjoys in the simulation. It is not clear what the causality is between the amount of fun and immersiveness. It could be that being immersive increases the fun of learning, or that the topic is intrinsically interesting to the player and contributes therefore to the immersiveness. Both causal directions are also possible, depending on the particular player.

Learning Fun versus Immersiveness				
Immersiveness	*n*	*R^2*	*Pearson Correlation*	*Significance*
Learning Fun	30	0,352	0,593	0,001

Table 16 Results Hypothesis 3D - Correlation Learning Fun versus Immersiveness

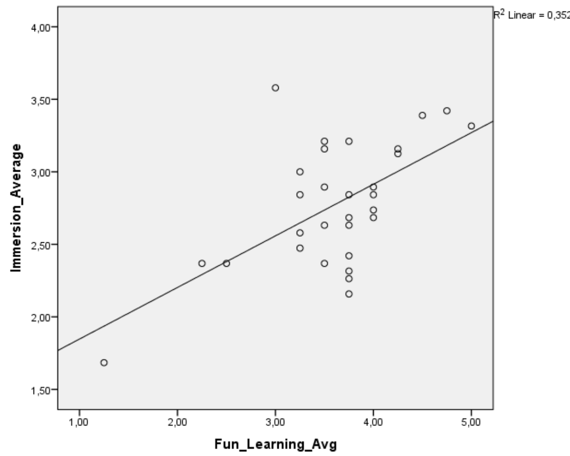

Figure 20 Correlation Learning Fun versus Immersiveness

Hypothesis 3E: *"Participants who have a good perception of the balance between the in-game reward structure and the difficulty of achieving this, enjoy a higher immersiveness than participants with an off-balance perception."*

There can be found no proof in the data supporting this hypothesis. All four trade-off indicators show an even spread over the immersion levels, indication that immersiveness of the player experiences no influence of the trade-off perception. In addition, the difficulty-reward trade-off deviates evenly on the score card on all four indicators. Meaning that there is no consistent opinion among participants that the simulation tasks are too easy, too hard or that the trade-off is sufficient.

Conclusion Immersiveness

The findings of this paragraph are that:

a) Immersiveness is correlated with the valuation of the simulation's design (Pearson Correlation coefficient = 0,53 & R^2=28%)

b) There is no observed distinction between different gaming backgrounds (both academic gaming background and business game experience) and the level of the participants' immersiveness. This means that players will experience a similar amount of immersiveness whilst playing the simulation, when only their gaming background differs.

c) Immersiveness is strongly correlated with the level of learning fun the participant experience. This indicates that if one is attracted to learn about the subject, he or she will be more likely to dive into the simulation than vice versa.
d) There is no consistency among players if they value the simulation too easy, too hard or that the difficulty – reward structure is in balance. The feedback is almost uniformly distributed.
e) In contrast to literature (Alvarez, 2008), there is no observation that there is a relation between the valuation of the simulation's difficulty – reward structure and the experience immersiveness.

5.1.5 Learning Efficacy

The main hypotheses for this research is about the final learning efficacy, being:

> Hypothesis: *"After playing Hazard Recognition, players experience competence improvement for the detection and assessment of hazards".*

Based on the assumption that the learning efficacy is built up from two constructs: the self-perceived competences and the objective performance, the conclusion can be drawn that with the finite amount of data available the hypothesis can be confirmed partly. The average improvements per construct, measured on a case by case basis, are shown in Table 17. The results teach us that learning efficacy is only achieved in the self-perceived competences, attitude towards hazards and self-perceived knowledge.

Quantitative Results Learning Efficacy

	n	*Mean Pre*	*Mean Post*	*Improvement*	*% Change*	*Significance**
Attitude	84	3,65	3,87	+0,22	+6%	0,000
Detection	80	3,83	3,88	+0,05	+1,3%	0,532
Recognition	83	3,63	3,82	+0,18	+5%	0,017
Responding	84	3,74	3,95	+0,21	+5,6%	0,006
Control	52	3,88	4,01	+0,13	+3,3%	0,107
Oversight	52	3,58	3,80	+0,23	+6,4%	0,003
SP Knowledge	52	3,50	3,82	+0,32	+9,1%	0,001
10 Hazards	52	8,71	8,71	+0,00	+0%	1,000
Procedural	41	0,73	0,78	+0,05	+6,8%	0,095

***Based on a 95% confidence interval of the difference | Paired Sample T-test**

Table 17 Main Hypothesis - Quantitative Results Learning Efficacy

Because the measurements shown in Table 17 are on a case by case basis, the measured learning efficacy is the improvement on the same scale. Because the non-industry participants were not assessed on objective performances before the simulation, their scores are excluded of these results. The reason for exclusion is that assessing laymen on procedural or knowledge topics, results are sketchy since they are not expected nor required to have knowledge or skills about the subject. When the objective performance results of the laymen after their participation are taken into account, shown in Table 18, we observe that this group scored relatively good. There are significant differences in scores between groups, but if we consider the fact that on commencement of the workshop they had absolutely no knowledge about hazards in chemical-plant environments, we can conclude that the simulation achieved significant learning amongst this group.

Therefore, players with or without a background in the oil and gas industry experience learning efficacy on the subject of hazard detection and assessment after playing the simulation Hazard Recognition.

Quantitative Results Learning Efficacy

	Oil & Gas Professional	*n*	*Mean*	*Difference (Significant: y/n)*	*Significance*
SP Knowledge Post	Yes	55	3,80	0,49 (Yes)	0,005
	No	30	3,31		
10 Hazards Post	Yes	54	8,67	2,59 (Yes)	0,000
	No	26	6,08		
Procedural Post	Yes	50	0,78	0,022 (No)	0,552
	No	26	0,75		

***Based on a 95% confidence interval of the difference | Independent Sample T-test**

Table 18 Main Hypothesis - Quantitative Results after Simulation

Results for all Hypotheses		
True | Not True | Partly True		**Hypothesis**
Partly True	0A	*"There is a relationship between the player's self-perceived competences and the player's performance on an objective scale."*
True	1A	*"Playing the simulation will cause a shift among all participants in their attitude towards hazards towards a more risk averse attitude."*
Partly True	1B	*"There is a positive relation between the attitude towards hazards measured before playing the game and the participants self-perceived competences measured both before and after the game."*
Partly True	1C	*"There is no difference in the increase in self-perceived competence between participants with or without a background in the oil & gas industry."*
Not True	1D	*"Participants with a background in the oil & gas industry have an inverse relationship between years of experience in the industry and their level of self-perceived competences (both before as after the simulation)."*
Not True	1E	*"Individual factors such as sex, nationality and highest level of education have no effect on the level of self-perceived competences and the amount of increase of the self-perceived competences."*
Not True	1F	*"There is a positive relation between the level of immersiveness of the participant and the increase of their self-perceived competences."*
Partly True	2A	*"There is a positive relation between the attitude towards hazard measured before playing the game and the participants objective performance measured both before and after the game."*
True	2B	*"Participants with a background in the oil & gas industry have a better objective performance than participants lacking this background."*
Not True	2C	*"There is a positive relation between the amount of years a participant has worked in the oil & gas industry and their objective performance."*
Partly True	2D	*"Individual factors such as sex, nationality and highest level of education have no effect on the objective performance nor on the objective learning efficacy."*
Not True	2E	*"Participants that experience a higher degree of enjoinment of learning about hazards have a higher objective performance than participants who enjoined this less."*
Not True	2F	*"There is a positive relation between the level of immersiveness of the participant and the objective performance."*
True	3A	*"Participants who have a higher regard of the simulation's design enjoy a higher degree of immersiveness."*
Not True	3B	*"Participants without a gaming background will enjoy an average feeling of immersiveness, participants with a gaming background will enjoy either an above average immersiveness due to their experience in the use of games or an beneath average due to their critical viewing of the simulation."*
Invalid Results	3C	*"The interference of structural bugs will have a negative impact on the participants level of enjoyment playing the game."*
True	3D	*"There is a positive relationship between the participant's level of learning fun and the level of immersiveness the participant enjoys."*
Not True	3E	*"Participants who have a good perception of the balance between the in-game reward structure and the difficulty of achieving this, enjoy a higher immersiveness than participants with an off-balance perception."*

Table 19 Results of the tested Hypotheses

5.2 Experiment Conduct Conclusion

The data for this study has been arrived with the means of four workshops. Two workshops were hosted within Shell, one at Accenture and one at Delft University of Technology. Resulting in a total experimental population of 91 people, whereof 60 people have a background in the oil and gas industry and 31 do not. Hence, the intention of an experimental group balanced between industry and non-industry was achieved. The total group consisted of people with 22 different nationalities and was for 85% male (and 15% female). The educational background of the participants was gravitated amongst BSc and MSc (together 80%). The data that has been gathered in these workshops and among these participants contributed to the answering of the following question:

"Does Hazard Recognition achieves learning efficacy and what factors contribute hereto?"

Yes, Hazard Recognition achieves learning efficacy within the experimental group, an overview of the results of the tested hypotheses is given in Table 19. We can conclude that learning efficacy is achieved because according to the measurements players experienced a movement through the quadrants of Burch's Competence Model. The observation is that novices tend to develop a conscious of their incompetence to detect hazards on worksites. In addition, novices' scores on the objective performance show that they experience learning after playing Hazard Recognition. For the professionals, this was not observed. A key aspect can be that the assessments were too easy and that therefore a saturation effect occurred. Future studies should develop assessments that are more meaningful in a professional context.

In addition the attitude of all players shifted to a more risk averse attitude, this means that Hazard Recognition makes people aware of dangers and participants observably convert this experience in a behavioral change. Which is proof that Hazard Recognition as a serious game does not only achieves learning efficacy but is also capable in changing behavior.

Also the laymen group enjoyed to learn about the topic of oil and gas industry specific hazards and mitigation procedures with the use of the simulation Hazard Recognition, indicating that this kind of simulation is not only effective but also an enjoyable literacy for the training of novices.

The factors that have an observable contribution to the learning efficacy are as follows. Since the learning efficacy is built up from two components, both components have been looked at separately and their relationship with the underlying factors. If we look at the underlying factors influencing the outcome of the <u>self-perceived competences</u> we see that participants with BSc background value their competences higher than participants with a MSc degree. The objective performance is influenced by the initial attitude towards hazards (positive relationship).

In this study's observation there has not been found a relationship between immersiveness and learning efficacy or in-game performance. However, literature suggests that these are related (Schooley, Moore, Schadler, & Catino, 2008). The only relations that have bee observed are that the level of immersion player's experience is positively correlated with the player's value of the simulation's design and also with the amount of learning fun they experienced.

6 Conclusion and Discussion

This study has been committed to the development of an evaluation method that can operate in the Comprehensive Evaluation Framework of Mayer et al. (2012) and which can evaluate a serious game's learning efficacy. In addition, this method has been applied to the evaluation of the learning efficacy of Hazard Recognition, a serious game developed by Delft University of Technology for Royal Dutch Shell NV. Therefore, this study's research questions is:

> *"What is an appropriate method, which can operate in the Comprehensive Evaluation Framework (Mayer et al., 2012), for the evaluation of a serious game's learning efficacy and the contributing factors?"*

In this research an appropriate method to evaluate a serious game's learning efficacy, within the context of Mayer's (2012) Comprehensive Evaluation Framework, has been laid down

A serious game's learning efficacy is defined as the ability of the learning intervention to achieve its learning objectives. This can be mapped by evaluating students exposed to a serious game and measure their movement through Burch's Competence Model (Burch, 1970s).

The method that has been proposed in this research is to map the students' journey through the quadrants of the Competence Model by the measurement of the relevant competences of the specific serious game in the dimensions of the two axes of the Competence Model (conscious and competence); self-perceived competences (conscious) and objective performance (competence).

The Kolb learning Cycle (A. Y Kolb & D. Kolb, 2008; D. Kolb, 1984) can best conceptualize the exposure of students towards a serious game and the learning process that accompanies this experience. In this cycle multiple cognitive tasks are required, as they are described in the Interactive Cognitive Complexion Model (ICC Model) (Tennyson & Breuer, 1997, 2002). In order to have a balanced measurement, and thus a balanced and more valid result, researcher should try to balance their measurements on all cognitive capacities as they are described in the ICC Model. In addition, it is valuable to understand in which way the serious game facilitates the four stages of the Kolb cycle. If this is understood it helps to select where emphasis in measurement should be.

Mayer's framework prescribes the five individual factors that could have moderating effects on the learning efficacy, i.e. underlying factors that contribute to the learning efficacy of the serious game. Depending of the scope of the evaluation study, the evaluation framework should be operationalized accordingly and an experimental group of participants should be selected. It is valuable to have a clear distinction in competences between two groups in the total experimental group; novices and experienced. This distinction helps to understand the initial position of the students in the Competence Model and functions as a reference line.

Within the theoretical concepts of Kolb, Burch and Tennyson & Breuer, the developed evaluation method prescribes a quasi-experimental approach with two observations (pre and post playing) for the gathering of data.

A closing part of the methodology is the formulation of hypotheses about the learning efficacy related concepts and underlying factors, subsequently these can be tested by the use of a selection of measurement tools. In this research there have been developed various constructs for the testing of a set of hypotheses that functioned as research tools for this research's case study; the learning efficacy and underlying factors of Hazard Recognition. This case study proved that the developed methodology is very capable to derive valuable knowledge on the learning efficacy of a serious game

Thus, there has been developed and successfully tested a quasi-experimental method that operates in Mayer's Comprehensive Evaluation Framework that is able to assess a serious game's learning efficacy and underlying, contributing factors. In addition, the conceptual model that has been laid down (combination of Burch, Kolb, Tennyson & Breuer) can function as a general model for the evaluation of serious games.

6.1 Discussion

6.1.1 Future Research

This research was the first step in the development of an evaluation method for a serious game's learning efficacy. There has been derived much knowledge and a valuable conceptual model that is able to assess a serious game's learning efficacy. The conceptual model that describes the learning efficacy and the relation to Kolb's Learning Cycle and the balancing of assessments (ICC Model) can function as a general model for the evaluation of serious games. However, the application of this conceptual model was very specific and can be extended towards other types of serious games. Since this model is only tested on a virtual 3D single player simulation.

In an interview with Rutger Deenen (Accenture) (Appendix III) it became clear that it would be very interesting for academia and in particular for Accenture if this method could be extended for the evaluation of serious games in a physical multiplayer environment. The learning objectives of these games are also more focused on soft skills and behavioral change, an extension of the current model towards an method that can assess a serious game's efficacy on these subjects should be subject of a future study and can contribute enormously to the existing knowledge base on game evaluation.

In addition, this study was mainly focused on the conceptual model of learning efficacy and its assessment. This resulted in a weak measurement tool for objective performances. One of the merits of this study is that it is very important that the objective assessment is relevant to the player and takes place in their world of reference. This means that the Shell professionals that were assessed in this research should have been engaged with a more difficult and sensible assessment with more distinctive power between competence levels of the different participants. Now this was not the case. So a major contribution for future studies would be a more sensible objective performance measurement tool.

Also, for the development of measurement tools that are able to specifically measure learning in different stages of the Kolb cycle it is necessary that the ICC Model is studied and coupled to the Kolb Cycle. This allows future researchers to target specific cognitive capacities of the student, resulting in more specific knowledge about the efficacy of the serious game.

6.1.2 Discussion on Findings

The observation of the results left many hypotheses unconfirmed or did not show the expected relationship between variables. Relationships between competences (self-perceived or objective) and underlying factors were not always found, even when such relationships were expected through reasoning, suggested in literature or a combination of both. In general the relationships that have been found have a low goodness of fit (R^2), which only on occasion exceeds 0,200. This is low and usually unacceptable. The reason for such low figure can be twofold. First, because the measurement tools lack enough distinctive power. Second, the amount of data might not have been enough or were gathered in a too homogeneous or a too deviated heterogeneous pool of players.

6.1.3 Discussion on the Research Method

The findings of this study rely heavily on the evaluation method that has been used. The method can be broken down into two concepts: 1) the theoretical concepts and evaluation framework, and 2) the designed measurement tools, operationalization and data gathering. During this study there has been gathered knowledge on the application of both concepts that will be discussed in this paragraph. This knowledge should contribute to the quality of future evaluation studies on serious games and simulations.

Theoretical Concepts

When reflecting on the theoretical concepts that have been used, we can conclude that they offer a solid base for this research. They facilitate all the important steps in order to discover the learning efficacy of Hazard Recognition based on the Comprehensive Evaluation Framework (I. Mayer et al., 2012a). The steps reflect the definition of what exactly should be investigated and how this is facilitated in the simulation:

1) What is the theory behind the learning objective of Hazard Recognition? Kowalski (Kowalski-Trakofler & Barret, 2003; Perdue et al., 1995)
2) What are competences and how do they develop? Competence Model (Burch, 1970s)
3) What is the theory behind the learning process of competences? Kolb Cycle (D. Kolb, 1984)
4) What cognitive capabilities does the student need in order to learn this? Interactive Cognitive Complexity Model (Tennyson & Breuer, 1997, 2002)
5) How does the simulation facilitate the learning process? Breakdown of the learning interventions.

These theoretical concepts define the input, process and output of the Comprehensive Evaluation Framework. In addition this framework defines very specifically the underlying factors. The authors opinion is that the merits of Mayer et al. (2012) offer a strong directive of the concepts that are of importance in an evaluation study. The framework is flexible enough to operationalize the information flow (input, process and output), but also specific in the underlying personal variables that reflect the participants characteristics and could explain moderating effects. However, the framework does not make an distinction between self-perceived competences and real competences.

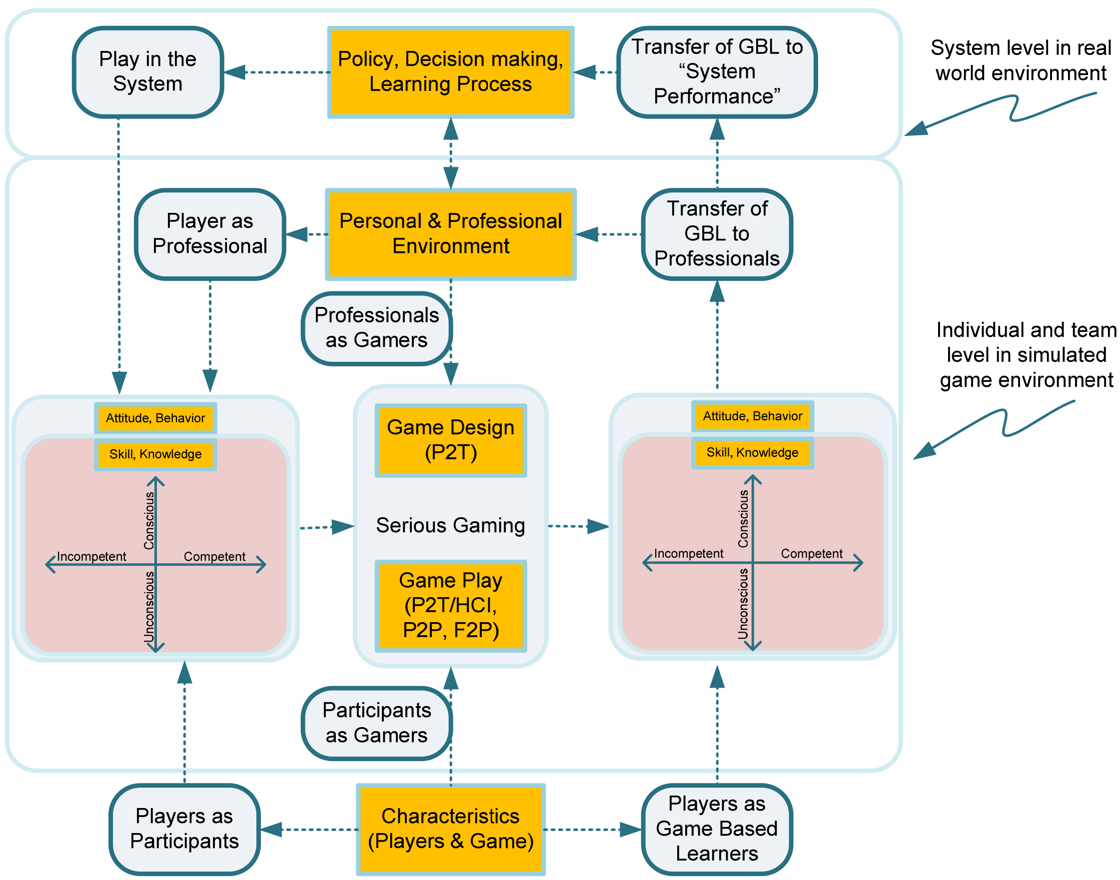

Figure 21 Comprehensive Evaluation Framework with Burch's Competence Model

This study shows that the combination of both the self-perception of competences as the real competences are important in order to understand the process of competence development. Hence, in order to understand the learning efficacy of a learning intervention both concepts are a necessity. Since serious games aim to change; behavior, skills or knowledge, i.e. facilitate a process of learning, the author holds the opinion that there should be room for this distinction in the framework in order to offer a guideline to map learning development better in future evaluations. A concept of this combination is shown in Figure 21.

Thus, the conceptual theoretical foundation of this research founds a strong base point for future research to the learning efficacy of serious games and in particular in the field of simulated hazardous environments. These merits are therefore highly recommendable for future evaluation studies.

Operationalization and Data Gathering

The result of the operationalization of the conceptual foundation and the final conduct of data gathering has more room for remarks. Regarding the measurement tool for the self-perceived competences, there were not many difficulties. This tool has been operationalized in a good manner and the response on the constructs were good. However, because the constructs were not mixed, the author suspects that some respondents were a little lazy during the filling in of the forms. Resulting in consecutive vertical lines with, for example, only 4's. However, this is only a suspicion and the response might be very valid. It is not possible to verify this.

Because the background variables and game experience constructs were built up with questions that had less similarities on face, they were filled in with more deviations on the Likert scales. This backs the earlier suspicion and is something future evaluation studies should keep in mind when developing their measurement tools.

Objective Performance

Regarding the measurements of the objective performance there is one major point of discussion. The objective performance was measured by the in-game performance (measured by in-game log files) and by a set of questions. The questions have one major flaw, they were too shallow for professionals. When analyzing the data, we can clearly see that the naming of hazards was exactly as high in the professional group before as after the simulation. This indicates a saturation point on the expected response, especially considered the fact that it is not a complicated task and that novices' scores yielded up to that of professionals. Regarding the procedural questions, they were also answered without much deviation.

The result of this problem is the lack of finding much relevant relationships. When the questions would have more distinctive power amongst professionals, there might have been more observed relationships between experience and performance. Or similar dependable variables would interact in such manner. This design choice was made because those tasks were not valued too easy for the laymen group and this approach would result in comparable data. A major recommendation for future evaluation studies is to develop a measurement tool that is able to assess professionals more in their own initial competence sphere.

In-Game Measurements

The collection of the in-game performances were confronted with two major problems. First, the laptops used at Accenture had administrator locks on the designated folder containing the log files. Hence, they were unable to retrieve. A similar problem occurred at the workshop at Delft University of Technology. Here the computers were connected to the internet during the game, which resulted in an automated process of sending the log files to the server and deleting them from the computer. The files were retrievable from the server, but without any identification to the computers. This made the files useless because they were unmatchable with the participants and their questionnaires.

A second complication is that log files are only stored when a participant finishes the whole level and manually submits the score. During the workshops it became clear that not everyone was able to finish the level completely, and needed to quit even after finishing 90% of the level. Information of those participants would have been of great value, but they were not stored. The same counts for the manual submitting of scores. This was unable to verify, but the author is confident that, despite of multiple stressing that it should be submitted, this has resulted in the perish of at least a few log files.

Hence, for future studies it would be highly recommendable to overcome these problems. Some could have been circumvented by the researcher (turning of internet, other destination folder), others are inherent characteristics that should be redesigned in future versions of the simulation (auto saving, save uncompleted levels).

Respondents

This has been stressed before, but deserves a particular paragraph. The participants that have been subject of this study were motivated to participate in totally different ways. The participants within Accenture participated on a totally voluntary basis. Whereas the participants at the Technical University of Delft participated as an obligatory part of their course (Game Design Project), although they choose this course because of their interest in serious gaming. We can say that these participants have an intrinsic motivation into playing serious games. The opposite was observed within the Shell, here the workshops were part of an obligatory training.

The observation during the simulation was that most participants were motivated to play and to achieve the in-game objectives. Only the response on the questionnaires was different within Shell than in the other voluntary groups. The Shell professionals handed in some empty or half answered questionnaires. This complacency can also be amongst other participants with a fully filled in questionnaire, but this is not easy to find out. The result would be that the gathered data is less reliable, consequently relationships that do exist are not/less well observable.

6.1.4 Future Adoption in to Shell

In previous chapters there has been created a thorough understanding of the simulation's capabilities and its reach into different target groups. The conclusion of this study is that the simulation achieves learning efficacy, no matter what the background of the player is. Also, the simulation can be considered as a complete learning method, incorporating all necessary features necessary for achieving a complete learning cycle. However, the adoption of the simulation into Shell's training curriculum is low. This can have several reasons that can exist on different levels in the organization, which could range from individual preferences of policy makers in the organization up to high level organizational strategies that don't align with preference of implementing the simulation into the organization. In this paragraph there will be discussed the experience of the author with regard to the adoption of Hazard Recognition into Shell and potential strategies that could smoothen this process.

Lewin's Force Field Analysis Model

The Force Field Analysis Model was developed for the purpose of explaining how organizational change processes work (see Figure 22). Although it is over 50 years old, the model remains one of the most widely respected ways of viewing this process according to recent reviews (McShane & Von Glinow, 2009). The original model is shown in Figure 22 in the blue box, for the sake of clarification two scenarios have been added, which will become clear later in this paragraph.

The model shows two types of forces. The upward aimed forces are the Driving Forces and represent the forces that push the organization to a new state of affairs. This can include new competitors, new technologies, evolving workforce expectations, or a host of other environmental changes. Even when a company is already leader in its market and when external forces for change aren't apparent, corporate leaders might choose to produce driving forces known as "divine discontent". Divine discontent means that leaders are continuously urging employees to strive for the highest standards and are constantly demanding to outshine the competition.

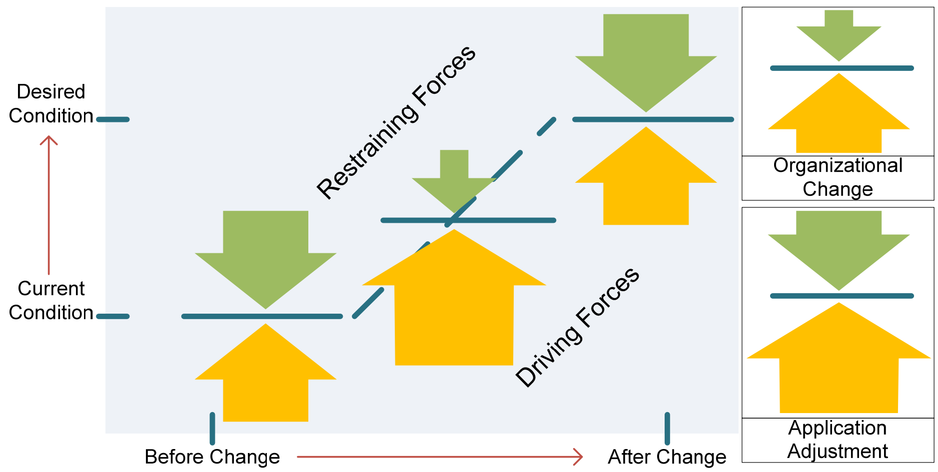

Figure 22 Lewin's Force Field Analysis Model - Strategic Addition

On the upside there are the Restraining Forces aiming downwards, which maintain the status quo of an organization. Most commonly these appear as employee behavior blocking the change process. Stability occurs when both the restraining as the driving forces are roughly in equilibrium, e.g. both forces are of approximately equal strength. If this equilibrium needs to be changed in favor of an upside movement, the driving force needs to outperform the restraining force. This process is called unfreezing. Unfreezing can happen by increasing the driving force, decreasing the restraining force, or by a combination of both. In the first option the driving force must be big enough to motivate change within the organization, however, change rarely occurs by increasing the driving force alone. But increasing the driving forces will increase an urgency for change within the organization, creating a pull movement of the change into the organization. When the driving force remains in its original condition, the organization needs to change in order to unfreeze the status quo (McShane & Von Glinow, 2009).

The problem Shell faces regarding the adoption of Hazard Recognition can be defined as:

"There is unnecessary hesitation within Shell to adopt Hazard Recognition into its training curriculum"

Because the scope of this study is solely about the characteristics and potential of the simulation, the following approach will be applied. The restraining force remains the organization's behavior towards the adoption process and will be considered as a given and unchangeable state. In this case the attitude within the Shell organization towards adopting new training literacies. Thus, there will not be advised upon an organizational change strategy for reducing the restraining force. Hence, the restraining forces will be identified, consequently a tailor made approach can be developed in order to get Hazard Recognition adopted into Shell.

Resistance Towards New Technology

Within the definition of Lewin, Hazard Recognition can be seen as a new technology that has the potential to push the status quo to a higher level. However, the technology's introduction hasn't caused a sense of urgency within the organization, at least not widespread enough. This lack of implementation urgency could have different origins in the organization. The organizational origin, refers to a possible common understanding within the organization (or at least amongst the important stakeholders) that Shell does not need or need to want an addition to its existing training literacies in the form of a serious simulation like Hazard Recognition.

The resistance that exists within the organization relies heavily on two components. These components are derived from the author's experience during this research and more importantly they are derived from experience of a senior expert within Accenture (interview transcript in Appendix III). The components are a knowledge barrier and an acknowledgement barrier.

The knowledge barrier relies steadily on the misunderstanding of stakeholders regarding the potential of serious games, or the process of implementation of serious games in an organization that is different than current methods and therefore scary to stakeholders.

The acknowledgement barrier represents the perception (and the consequences thereof) of stakeholders who recognize that serious gaming as a good addition to current training methods would indicate that current methods are not maintaining the highest safety standards. Hence, perception of acknowledging that the organization operates not as safe as it could be operating with the use of extra training capabilities.

Organizational Strategy Alignment Issues

In addition, there are alignment issues that conflict with current strategies that would block a smooth adoption of Hazard Recognition. Shell's valuation of the serious gaming market is one that the market is not mature, resulting in the decision not to buy or develop serious games like Hazard Recognition now. The decision not to develop serious games is nested in the perception that game development is not one of Shell's core businesses and that serious games are commodities that are not tailor made solutions, which is a misconception. Serious games, especially the ones like Hazard Recognition, are per definition tailor made solutions that require the full commitment of the full organization.

This lights the second important alignment issue, the organizational support for serious games. Hazard Recognition is a learning tool, which derives its objectives from Human Resources, it is supported by IT and the merits of learning are materialized in the Business. At this moment there is no alignment in the ownership of simulations. This makes the responsibility sharing difficult and is therefore not happening. The consequence is a hesitative attitude in the organization that leads to unnecessary delays or even cancelation of implementation.

A related consequence to this issue is that the learning within the Shell organization will not be facilitated as it should be. Organizational supported learning works in three steps, which are shown in Figure 23. These three steps are aimed to the student and should be facilitated by the organization. Hence, the student should be motivated, then enabled and experience a support in the learning process by the organization. The absence of one of these facilitations will result in a sloppy learning process. The lack of ownership can result in the missing of at least one of the three facilitations (motivating, enabling, supporting), shared responsibilities should therefore be incorporated.

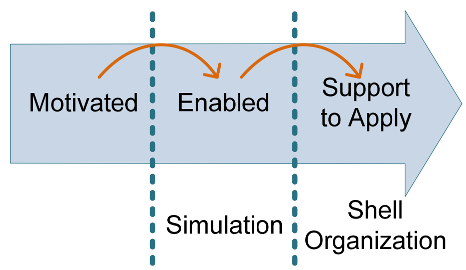

Figure 23 Organizational Learning Support

These shared responsibilities would be between the simulation objective setter (Human Resources), the facilitator (IT) and the beneficiary (Business). Ultimately a platform consisting of all three stakeholders in the organization would be best able to motivated, enable and facilitate future application of serious gaming without the risk that essential facilities lack.

Organizational Performance Model

The existing absence of a shared facilitation throughout the organization for serious game learning methods can be identified by the use of the Organization Performance Modell used within Accenture (Appendix III). Where the conceptualization of Organization Learning Support (Figure 23) visualizes the steps that a student should be facilitated in, in order to successfully enter a learning process, the Organizational Performance Model (Figure 24) identifies the facilities that facilitate these steps and in this case where the responsibilities should be shared.

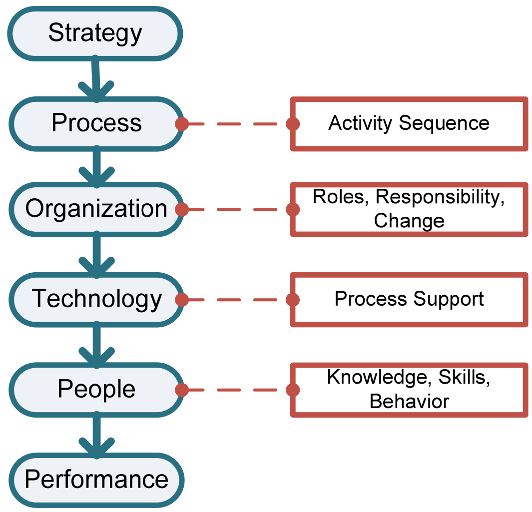

Figure 24 Accenture Organization Performance Model

The model visualizes the necessity of organizational and technical support for learning processes to be successful in the achievement of their objectives and consequently these learning processes will improve the employees' knowledge, skills and/or behavior.

The Barriers to Overcome

The observations of the barriers that block adoption of Hazard Recognition, and simulations alike, into Shell are mainly shaped by opinions of stakeholders within the organization, rather than by inabilities driven by technical solutions or funds. This is in line with the observation of Lewin, that in most cases the restraining forces are shaped by the unwillingness of stakeholders to adopt or change (McShane & Von Glinow, 2009). The opinions of stakeholders within Shell are formed on the misperception, the misunderstanding or the lack of knowledge of serious games and their use.

In response to knowledge barriers, new institutions come into existence which progressively lower those barriers, and make it easier for firms to adopt and use the technology without extensive in-house expertise. Service bureaus, consultants, and simplification of the technology are examples. As knowledge barriers are lowered, diffusion speeds up, and one observes a transition from an early pattern in which the new technology is typically obtained as a service to a later pattern of in-house provision of the technology (Attewell, 1992).

Hence, Shell should open the organization towards the new capabilities, be open to learn about serious games and shouldn't be afraid to implement new institutions in the organization necessary for the support and application of the newly acquired serious game(s).

Literature

Alvarez, J. (2008). *Serious games*. Montpelier, France: IDATE Consulting & Research. Retrieved from www.idate.org

Amory, A., & Seagram, R. (2003). Educational game models: conceptualization and evaluation. *South African Journal of Higher Education*, *17*(2), 206 – 217.

Attewell, P. (1992). Technology Diffusion and Organizational Learning: The Case of Business Computing. *Organization Science* , *3* (1), 1–19. doi:10.1287/orsc.3.1.1

Bouffard-Bouchard, T. (1990). Influence of Self-Efficacy on Performance in a Cognitive Task. *The Journal of Social Psychology*, *130*(3), 353–363. doi:10.1080/00224545.1990.9924591

Boyle, E. a., Connolly, T. M., & Hainey, T. (2011). The role of psychology in understanding the impacht of computer games. *Entertainment Computing*, *2*, 69–74.

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

Burch, N. (n.d.). The Four Stages for Learning Any New Skill. CA, USA: Gordon Training International. Retrieved from http://www.businessballs.com/consciouscompetencelearningmodel.htm#origins of conscious competence learning model

Connolly, T. M., Boyle, E. a., MacArthur, E., Hainey, T., & Boyle, J. (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

Faria, a. J., Hutchinson, D., Wellington, W. J., & Gold, S. (2008). Developments in Business Gaming: A Review of the Past 40 Years. *Simulation & Gaming*, *40*(4), 464–487. doi:10.1177/1046878108327585

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

HARC, & Haut, R. (n.d.). EFD Virtual Rig. Retrieved March 17, 2013, from http://www.efdvirtualsite.org/

Hainey, T., Connolly, T., Stansfield, M., & Boyle, E. a. (2011). The difference in motivations of online game players and offline game players: a combined analysis of three studies at higher education level. *Computers & Education*, *57*, 2197–2211.

Haller, M., Kurka, G., Volkert, J., & Wagner, R. (2001). omVR - A Safety Training System for a Virtual Refinery. FAW - Institute for Applied Knowledge Processing.

Harteveld, C., Guimaraes, R., Mayer, I. S., & Bidarra, R. (2009). Balancing Play, Meaning and Reality: The Design Philosophy of LEVEE PATROLLER. *Simulation & Gaming*, *41*(3), 316–340. doi:10.1177/1046878108331237

Harteveld, Casper. (2010). Triadic Game Evaluation: A framework for assessing games with a serious purpose. Delft: Delft University of Technology.

Harteveld, Casper, Guimaraes, R., Mayer, I., Bidarra, R., & Guimarães, R. (2007). Balancing pedagogy, game and reality components within a unique serious game for training levee inspection. *Edutainment* (p. 12). Hong Kong.

Howland, G. (1999). Game Design: Immersiveness. Retrieved from http://www.gamedev.net/page/resources/_/creative/game-design/game-design-immersiveness

Kirriemuir, J., & McFarlane, A. (2004). Literature review in games and learning. Bristol: Nesta Futurelab.

Kolb, A. Y., & Kolb, D. a. (2008). The Learning Way: Meta-cognitive Aspects of Experiential Learning. *Simulation & Gaming*, *40*(3), 297–327. doi:10.1177/1046878108325713

Kolb, D. (1984). *Experiential Learning: Experience As The Source Of Learning And Development*. Englewood Cliffs, NJ, USA: Prentice Hall.

Kowalski-Trakofler, K., & Barret, E. (2003). The concept of degraded images applied to hazard recognition training in mining for reduction of lost-time injuries. *Journal of Safety Research*, *34*, 515–525.

Leutner, D. (1993). Guided discovery learning with computer-based simulation games: effects of adaptive and non-adaptive instrucutional support. *Learning & Instruction*, *3*, 113–132.

Luszczynska, A., & Schwarzer, R. (2005). Social cognitive theory. In M. Conner & P. Norman (Eds.). *Predicting health behaviour* (pp. 127–169). Buckingham, England: Open University Press.

Manolis, C., Burns, D. J., Assudani, R., & Chinta, R. (2013). Assessing experiential learning styles: A methodological reconstruction and validation of the Kolb Learning Style Inventory. *Learning and Individual Differences*, *23*, 44–52. doi:10.1016/j.lindif.2012.10.009

Mayer, I., Bekebrede, G., Harteveld, C., Lo, J., Ruijven, T. van, Warmelink, H., & Zhou, Q. (2012a). The Research and Evaluation of Serious Games: Towards a comprehensive methodology.

Mayer, I., Bekebrede, G., Harteveld, C., Lo, J., Ruijven, T. van, Warmelink, H., & Zhou, Q. (2012b). The Research and Evaluation of Serious Games: Towards a comprehensive methodology.

Mayer, I. S., Meijer, S., Verbraeck, A., Poelman, R., Warmelink, H., Harteveld, C., Veen, W., et al. (2009). Towards Better HSE Performance in Royal Dutch Shell. Rijswijk: Shell, TU Delft.

McShane, S. L., & Von Glinow, M. A. (2009). *Organizational behavior : emerging knowledge and practice for the real world*. Boston, MA, USA: McGraw-Hill.

Millsap, R., & Maydeu-Olivares, A. (2009). *The SAGE Handbook of Quantitative Methods in Psychology*. SAGE Publications Ltd.

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, *97*(1), 117–128.

Ormrod, J. E. (2006). *Educational psychology: Developing learners (5th ed.)*. Upper Saddle River, NJ, USA: Pearson/Merrill Prentice Hall.

Perdue, C., Kowalski-Trakofler, K., & Barret, E. (1995). Hazard Recognition in Mining: A Psychological Perspective. Washington DC: United States Department of the Interior.

Schooley, C., Moore, C., Schadler, T., & Catino, S. (2008). *For Stickier Learning, Try a Dose Of Serious Gaming*. Camebridge, MA, USA.

Schuurink, E., Houtkamp, J., & Toet, A. (2008). Engagement and EMG in serious gaming: Experimenting with sound and dynamics in the Levee Patroller training game. (P. Markopoulos, Ed.)*Fun and Games*, *Lecture No*, 139–149. doi:10.1007/978-3-540-88322-7_14

Shell. (2009). Project Initiation Note (PIN Document).

Shell. (2013a). Safety Record. Retrieved March 17, 2013, from http://www.shell.com/global/aboutshell/our-strategy/major-projects-2/pearl/safety-record.html

Shell. (2013b). The Shell Eastern Petrochemicals Complex. Retrieved March 17, 2013, from http://www.shell.com/chemicals/aboutshell/media-centre/factsheets/shell-eastern-petrochemicals-complex.html

Spek, E. D. V. D., Wouters, P., & Oostendorp, H. V. (2011). Code Red : Triage Or COgnition-based DEsign Rules Enhancing Decisionmaking TRaining In A Game Environment, *42*(3), 441–455. doi:10.1111/j.1467-8535.2009.01021.x

Squelch, A. P. (2001). Virtual reality for mine safety training in South Africa. The South African Institute of Mining and Metallurgy.

Tate, D. L., Sibert, L., & King, T. (1997). Virtual environments for shipboard firefighting training. *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*, 61–68,. doi:10.1109/VRAIS.1997.583045

Tennyson, R. D., & Breuer, K. (1997). Psychological foundations for instructional design theory. *Instructional Design: International perspective, volume 1: Theory, research, and models*. Mahwah, NJ: Lawrence Erlbaum Associates.

Tennyson, R. D., & Breuer, K. (2002). Improving problem solving and creativity through use of complex dynamic simulations. *Computers in Human Behavior*, *18*, 650–668.

Tennyson, R. D., & Jorczak, R. L. (2008). A conceptual framework for the empirical study of instructional games. *Computer Games and Team and Individual Learning*. Elsevier Ltd.

Vora, J., Nair, S., Gramopadhye, A. K., Duchowski, A. T., Melloy, B. J., & Kanki, B. (2002). Using virtual reality technology for aircraft visual inspection training: presence and comparison studies. *Applied ergonomics*, *33*(6), 559–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12507340

Warmelink, H., Meijer, S., Mayer, I., & Verbraeck, A. (2009). Introducing Serious Gaming in a Multinational : Experiences with the Supervisor Serious Game for HSE Training, *2009*.

Watson, P. C. (1963). On the failure to eliminate hypotheses - A second look. *Thinking and Reasoning*. Baltimore, MD, USA: Penguin Books.

Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*(4), 263–290. doi:10.1002/bdm.414

Williams, G. C., & Deci, E. L. (1996). Internalization of biopsychosocial values by medical students: A test of self-determination theory. *Journal of Personality and Social Psychology*, *70*, 767–779.

Williams, G. C., Freedman, Z. R., & Deci, E. L. (1998). Supporting autonomy to motivate glucose control in patients with diabetes. *Diabetes Care*, *21*, 1644–1651.

Wolfe, J. (1993). A history of business teaching games in English-speaking and post-socialist countries: The origination and diffusion of a management education and development technology. *Simulation & Gaming*, (24), 446 – 463.

van Wyk, E., & de Villiers, R. (2009). Virtual Reality Training Applications for the Mining Industry. Tshwane University of Technology.

Appendix IA

Hypotheses on Self-Perceived Competences and Objective Performance

With regard to the conscious competence matrix discussed in paragraph 0, the relationship between the players' objective performance and the players' self-perceived competences is of interest. This relationship could give insight in the amount of confidence that could be laid in a statement of a player's own competences. Thus, is the perceived competences scale a good method to get an insight in the objective performance of a person, within the context of hazard recognition of course. In addition, there will be looked at the total group and there will be made a distinction between oil & gas professionals and outsiders. Thus the hypotheses would be:

> Hypothesis 0A: *"There is a relationship between the player's self-perceived competences and the player's performance on an objective scale."*

Appendix IB

Hypotheses on Self-Perceived Competences

The self-perceived competences of the participants to detect hazards described by Kowalski (Kowalski-Trakofler & Barret, 2003; Perdue et al., 1995) is perceivably influenced by four factors show in Figure 25. Since this study is focused on the learning efficacy of Hazard Recognition the focus of the analysis will lie on the increase of the self-perceived competences and not on their individual values, off-course they are measured and analyzed in some cases but are not always relevant in others. The presumable explanations are given in the following hypotheses accompanied by a substantiation.

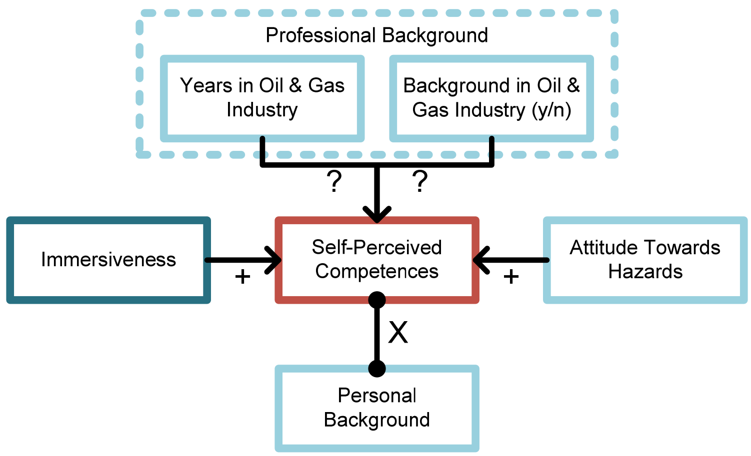

Figure 25 Hypotheses Relationships Self-Perceived Competences

Attitude Towards Hazards

Hypothesis 1A: *"Playing the simulation will cause a shift among all participants in their attitude towards hazards towards a more risk averse attitude."*

Considering the attitude towards hazards among all participants, it can be expected that playing Hazard Recognition will cause a shift towards a more risk averse attitude. This assumption is based on the fact that the plain observation of hazards will cause an awareness effect, which refreshes the mind on the consequences of hazards. This would imply that the attitude towards hazards would undergo a similar Kolb Cycle into the generalization of consequences as the other learning objectives.

Hypothesis 1B: *"There is a positive relation between the attitude towards hazards measured before playing the game and the participants self-perceived competences measured both before and after the game."*

Weber (Weber, Blais, & Betz, 2002) underpins that one's attitude towards risk is a common dominator on the perception of risk within a specific domain and the behavior within that domain. The assumption of hypothesis 1B is that a conscious attitude towards risk would imply that one is more alert on risks and because of that higher level of awareness one will perceive their own competences regarding the detection of risks/hazards higher. A consequence of this assumption is that when one's competence levels are initially higher on a 1 to 5 scale, the increase potential after the game is smaller. Therefore a lower increase should be expected as a result of this.

Professional Background

Hypothesis 1C: *"There is no difference in the increase in self-perceived competence between participants with or without a background in the oil & gas industry."*

Self-perceived competences are self-measured levels and are therefore subjective. Assuming that the increase in this perception level is similar amongst all participants means that conscious making develops similar amongst both groups.

Hypothesis 1D: *"Participants with a background in the oil & gas industry have an inverse relationship between years of experience in the industry and their level of self-perceived competences (both before and after the simulation)."*

Considering the logic set out in chapter 0 about the conscious-competence quadrant, the assumption is that participants with a professional background in the oil & gas industry fall per definition in the competence area. Within this group it can be expected that participants working a large number of years in the industry fall in the final area of the conscious-competence quadrant (Figure 10); unconscious-competent. Novices in the industry would therefore be conscious-competent. Table 3 is a summary of the distilled logic from that chapter and shows that according to this logic experienced professionals would perceive their competences lower than novices, presumably due to their greater amount of knowledge and the accompanying knowledge of what they do not know. Hence, an inverse relationship between self-perceived competences and years of experience should be witnessed within the results of the workshop.

Personal Background

Hypothesis 1E: *"Individual factors such as sex, nationality and highest level of education have no effect on the level of self-perceived competences and the amount of increase of the self-perceived competences."*

However age is expected to be correlated with years of experience and is therefore not independent, the main expectations is that self-perceived competences do not differ amongst participants based on sex, age, nationality or education level alone.

Immersiveness

Hypothesis 1F: *"There is a positive relation between the level of immersiveness of the participant and the increase of their self-perceived competences."*

Immersiveness is a factor that has a big influence on the gaming experience and the in-game achievements of the player (Alvarez, 2008; Howland, 1999; Schooley et al., 2008). Being highly immersive as a player often means that the player is absorbed to a level that the game matters above all else and are fully focused on the gaming experience with a sense of pure pleasure. With regard to Hazard Recognition, it may be expected that players with a high level of immersiveness have the feeling that their in-game activity was performed good. This would result in the perception of the player that their derived knowledge is high, a feeling more prominently present with highly immersed players than lower immersed players.

Appendix IC

Hypotheses on Objective Performance

The *objective performance* of the participants is measured with questions about hazard procedures and with the assessment of the participant's in-game score. The questions measure what the participants already knew and what was learned about hazards and procedures after playing the game, the in-game performance measures show how much hazards were detected on time. This means that the objective performance includes both the learning improvement of the participant and the in-game performance. Most of the factors influencing the *objective performance* are similar to the *self-perceived competences*, but the relationships are significantly different. Where the self-perceived competences are all about the participants own perception of his or her competences, the *objective performance* is measurable by a third party objectively. As a result initial capabilities and background have different effects.

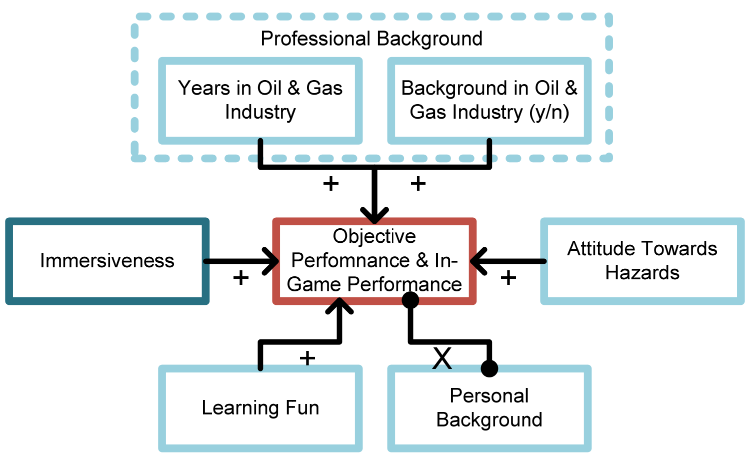

Figure 26 Hypotheses Relationships Objective Performance

Attitude Towards Hazards

Hypothesis 2A: *"There is a positive relation between the attitude towards hazards measured before playing the game and the participants objective performance measured both before and after the game."*

As previously argued, the attitude towards hazards is a strong factor on the perception of risk and the accompanied behavior. This was argued for the self-perceived competences of the participant, but this same statement counts even more so for the objective performance of the participant because of the direct influence of attitude towards risk on behavior in risk environments (Weber et al., 2002).

Professional Background

Hypothesis 2B: *"Participants with a background in the oil & gas industry have a better objective performance than participants lacking this background."*

Hypothesis 2C: *"There is a positive relation between the amount of years a participant has worked in the oil & gas industry and their objective performance."*

The expectation is that participants with a background in the oil & gas industry will perform better than participants due to the nature of their work experience and training history than participants without this experience. In addition it can be expected that the amount of years the participant has worked in the industry will have a positive influence on the objective performance.

Personal background

Hypothesis 2D: *"Individual factors such as sex, nationality and highest level of education have no effect on the objective performance nor on the objective learning efficacy."*

However age is expected to be correlated with years of experience and is therefore not independent, the main expectations is that the objective performance does not differ amongst participants based on sex, age, nationality or education level alone.

Learning Fun

Hypothesis 2E: *"Participants that experience a higher degree of enjoinment of learning about hazards have a higher objective performance than participants who enjoined this less."*

The amount of fun/interest players experience about the learning of hazards would indicate, amongst others, their willingness to learn about hazards. The assumption is that people who enjoy learning about this subject would perform better than players who do not. This assumption is funded by the common observation that people commonly do not achieve well on subjects they are not interested in.

Immersiveness

Hypothesis 2F: *"There is a positive relation between the level of immersiveness of the participant and the objective performance."*

As argued for the relationship between self-perceived competences and immersiveness, the same goes for the objective performance: a higher degree of immersiveness will be correlated with the objective performance of the player. The main assumption is that players who are immerged more, have a higher level of concentration and will therefore perform better both in the game and outside the game when it comes down to learning about hazards.

Appendix ID

Hypotheses on Immersiveness

Since immersiveness is considered as an influencing factor on the player's achievements within games(Schooley et al., 2008), there will be taken a closer look on the surroundings influencing immersiveness. The player's level of immersiveness can be measured independently with a validated questionnaire (Brockmyer et al., 2009). This could help us understand learning efficacy differences between players based on the level of immersiveness they experienced, but this is a passive measurement that does not reveal the game's ability to facilitate immersiveness among players. It only discovers if there is immersiveness and if so, how much. Immersiveness depends on a decomposition of two constructs: player's individual sensitivity to immersiveness and the game its facilitation by design. When evaluating Hazard Recognition, the players' general sensitivity to become immersed is a given, we do not control this nor are we aiming to improve this by training. However, the ability by design of the simulation to get players immersed is within the evaluation scope. In this study there have been identified five parameters, shown in Figure 27, that have influence or discriminate amongst players on the simulation its ability to create immersiveness.

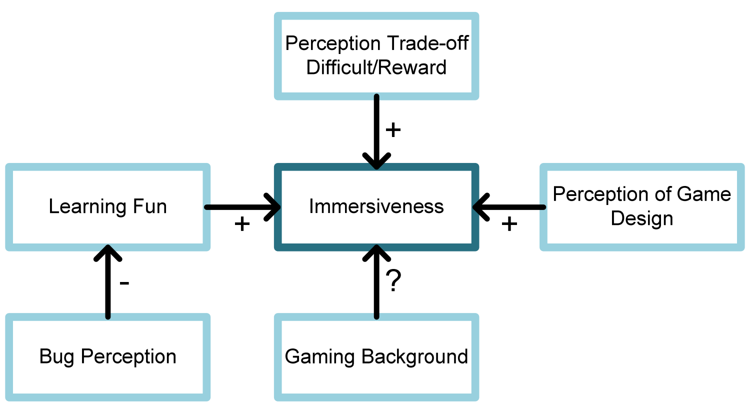

Figure 27 Hypotheses Relationships Immersiveness

Perception of Game Design

Hypothesis 3A: *"Participants who have a higher regard of the simulation's design enjoy a higher degree of immersiveness."*

Looking at the players' perception of Hazard Recognitions its design gives insight in two things. For one, it shows the overall perception and valuation of the game's design and functionality. This helps us understand how the game lands in its target group. Two, the game's design is the main construct delivering immersiveness. Therefore, when it is understood what the general perception of the design and functionality of the game is, it is possible to find a relation between the deliverable of the mechanics, the immersiveness and the perception of the quality of the mechanics. The assumption is that a high value of the game's design will be related with a higher immersiveness of the player.

Gaming Background

Hypothesis 3B: *"Participants without a gaming background will enjoy an average feeling of immersiveness, participants with a gaming background will enjoy either an above average immersiveness due to their experience in the use of games or an beneath average due to their critical viewing of the simulation."*

The hypothesis distincts two groups, players with or without a gaming background. The assumption is that the group of players without a gaming background will have their immersiveness levels closely together around the mean: called the *core*. The values above and below the *core* are inhabited by players with a background in gaming, because their experience will help them to get more immersed or their knowledge on games makes them more critical. There is no literature supporting this hypothesis.

Bug Perception and Fun

Hypothesis 3C: *"The interference of structural bugs will have a negative impact on the participants level of enjoyment playing the game."*

The occurrence of bugs whilst playing will perceivably have an effect on the simulation's experience (Casper Harteveld, 2010). The expected relation is that being exposed to bugs reduces directly the fun and immersiveness of the player because the bug will get the players out their *flow*.

Hypothesis 3D: *"There is a positive relationship between the participant's level of learning fun and the level of immersiveness the participant enjoys."*

The expectation is that players who are attracted to the subject of hazards and find it enjoyable to learn about the subject are more dived into the game because of their learning motif than players who find this subject not enjoyable.

Hypothesis 3E: *"Participants who have a good perception of the balance between the in-game reward structure and the difficulty of achieving this, enjoy a higher immersiveness than participants with an off-balance perception."*

Literature provides us with the knowledge that immersiveness partly depends on the in-game trade-offs, players make between difficulty of tasks and accompanying rewards (Alvarez, 2008). Also, the game's challenge is a high motivator for players to compete and continue playing (Hainey, Connolly, Stansfield, & Boyle, 2011). Within the scope of this study these trade-offs are particularly interesting, because the difficulty of tasks and the reward structure is the core of the learning content transfer. Researching this hypothesis would give insight in the overall quality of the task difficulty and in which extend this influences the gaming experience as a whole.

Appendix IIA

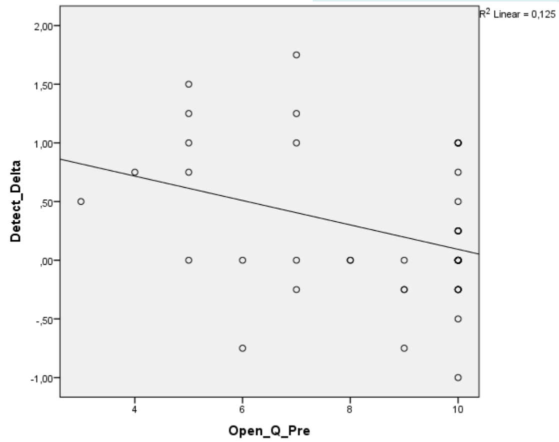

Figure 28 Detect Delta vs 10 Hazards Pre (Only Shell Participants)

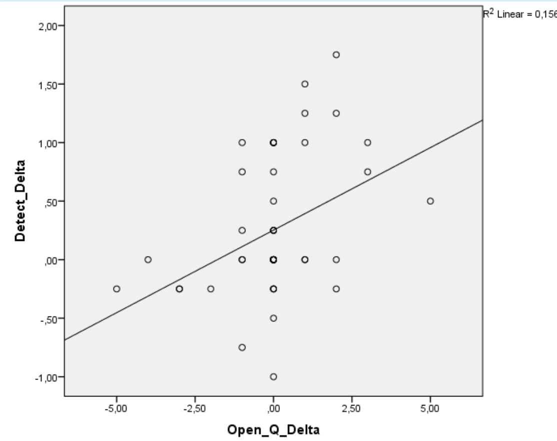

Figure 29 Detect Delta vs 10 Hazards Delta (Only Shell Participants)

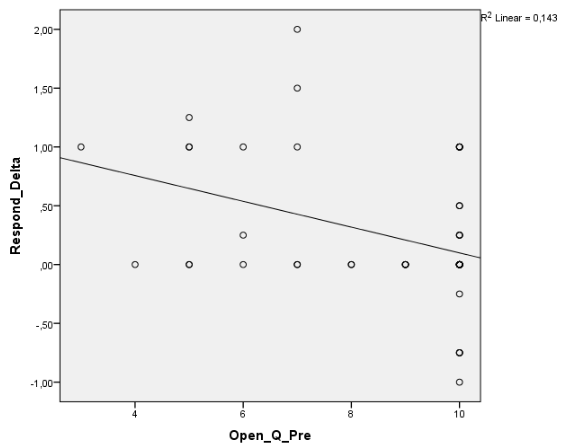

Figure 30 Respond Delta vs 10 Hazards Pre (Only Shell Participants)

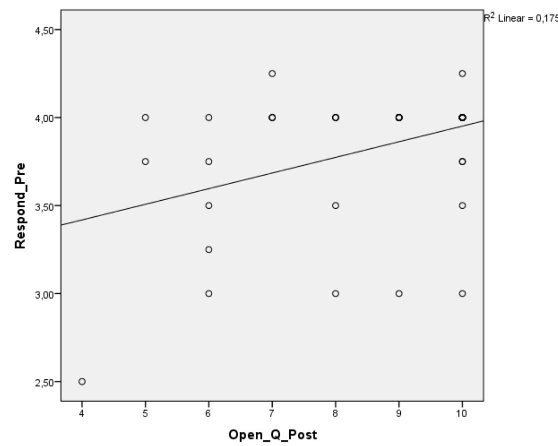

Figure 31 Respond Pre vs 10 Hazards Post (Only Shell Participants)

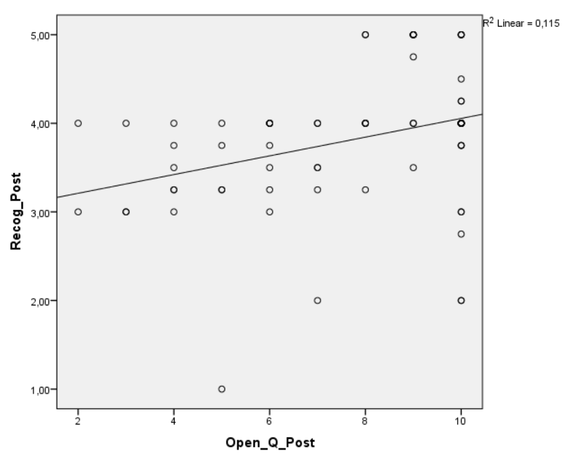

Figure 32 Recognition Post vs 10 Hazards Post (All Participants)

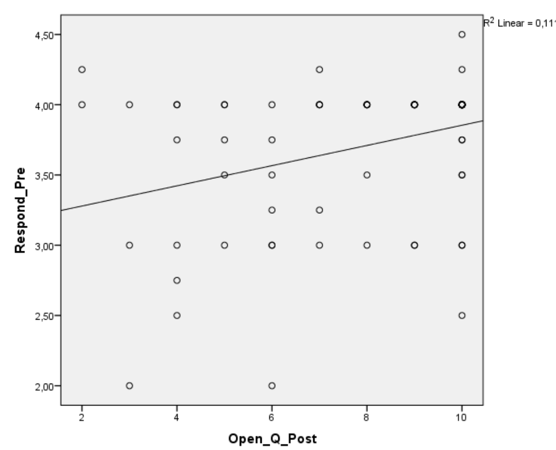

Figure 33 Respond Pre vs 10 Hazards Post (All Participants)

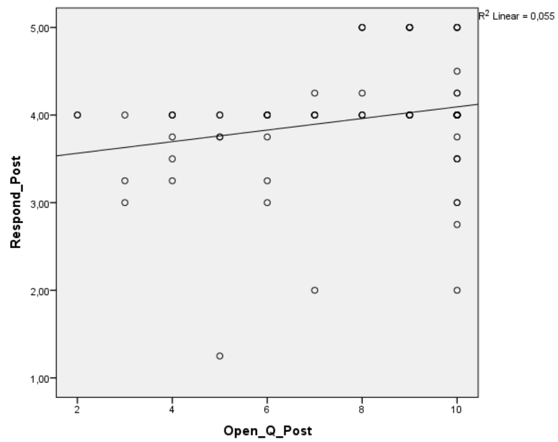

Figure 34 Respond Post vs 10 Hazards Post (All Participants)

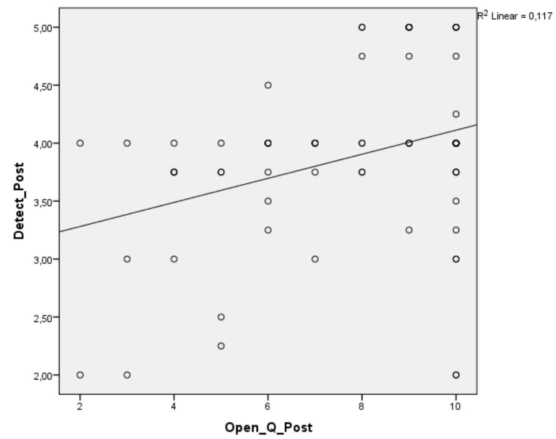

Figure 35 Detect Post vs 10 Hazards Post (All Participants)

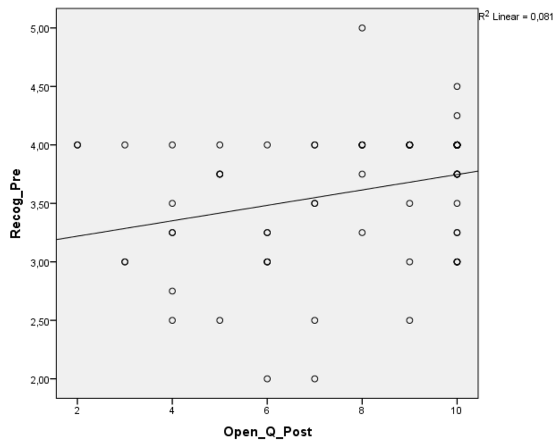

Figure 36 Recognize Post vs 10 Hazards Post (All Participants)

Appendix IIB

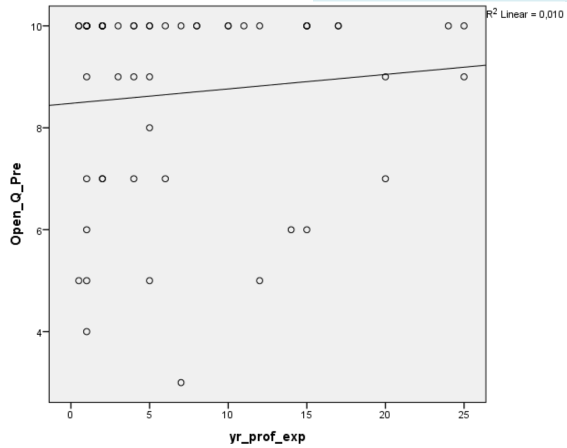

Figure 37 10 Hazards Pre versus Years of Professional Experience

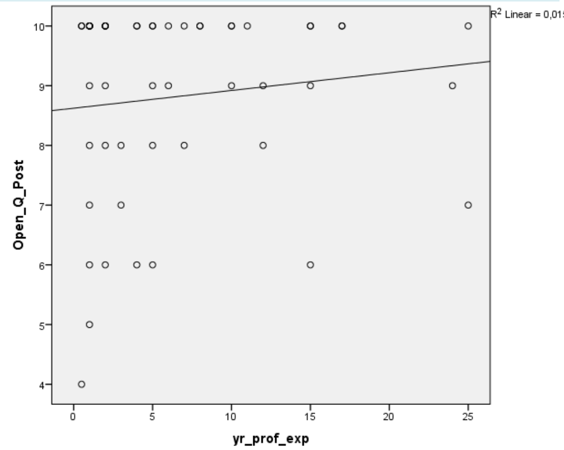

Figure 38 10 Hazards Post versus Years of Professional Experience

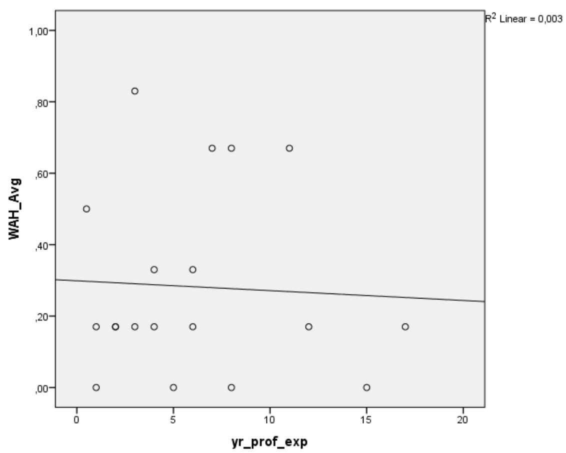

Figure 39 Score Level Working at Heights versus Years of Professional Experience

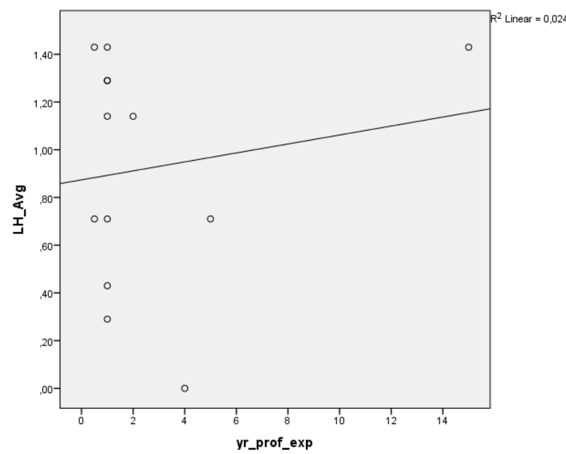

Figure 40 Score Level Lifting and Hoisting versus Years of Professional Experience

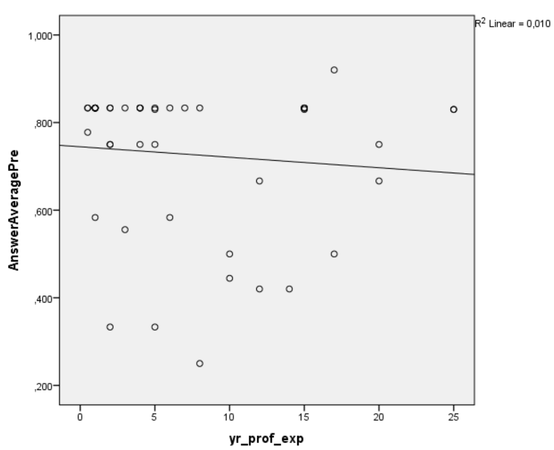

Figure 41 Procedural Questions Pre versus Years of Professional Experience

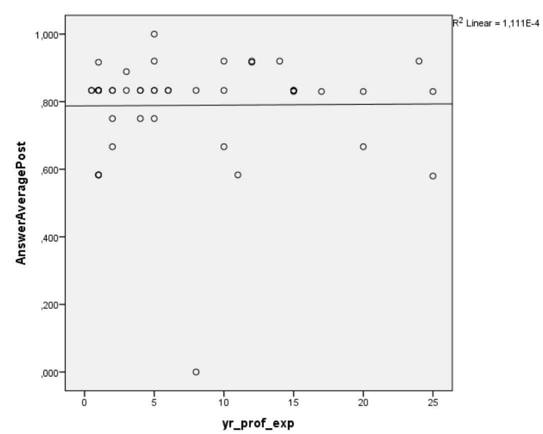

Figure 42 Procedural Questions Post versus Years of professional Experience

Appendix III

Interview Ivo Wenzler

The following statements have been derived from the interview on Tuesday 26 February 2013.

On the question what could be possible barriers that would resist implementation of a serious game within Shell, the following quotes have been noted:

- "It's not invented here"
- "Implementation takes time and effort, with the perception that the return on this investment is guaranteed"
- "Games are not considered to be serious, an issue of perception"
- "Games require different attention of an organization than traditional methods, which makes adopters hesitative"

On the question what perceptions towards serious gaming are common in organizations like Shell, the following quotes have been noted:

The perception is that…

- "Serious games are expensive"
- "Serious games are not really serious"
- "Games are a commodity and can be bought off the shelve, instead games are custom made and that requires money and sponsorship."
- "In-house development is scary, requires responsibility and the output is unclear"
- "A serious game does not deliver what the seller claims it delivers"
- "The in-house knowledge is too low in order to understand how to work with serious games"
- "Organizations don't think that their traditional way of working needs contribution from other methods in order to increase organization's performance"

Regarding the barrier to adopt serious games into the organization because of the lack of a sense of urgency, the interviewee said the following:

- "A company like Shell, which operates in an industry where safety is a core pillar of the business, companies will be reluctant to admit there are training methods available that can result in better safety. This would indicate that the company is not as safe as it potentially could be."

- Organizational responsibility and the consequences this have to the adoption process have de following barriers:
- "There is always the question of ownership when serious games like Hazard Recognition are implemented. There are three pillars facilitating the game, who owns it HR, Business, IT? The traditional organizational structure needs to be changed in order to facilitate this."

Organizational Performance model

Accenture uses the following model (Figure 43) in order to identify the essential facilities that are necessary to materialize a business strategy into preferred performance.

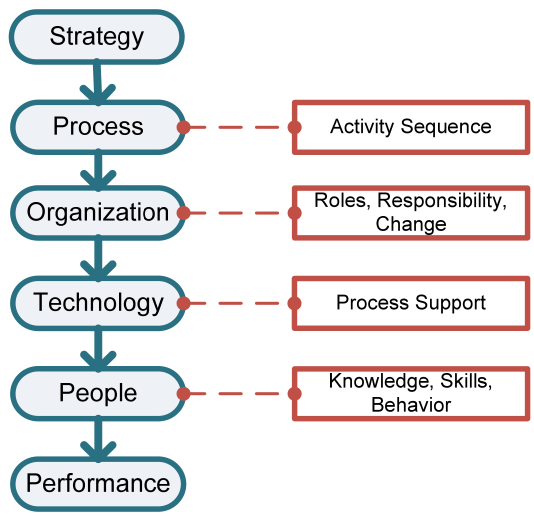

Figure 43 Organizational Performance Model (Accenture)

Organizational Learning Support

During the interview, dr. Wenzler visualized the following steps (Figure 44). He thinks that are a necessity in an organization in order to facilitate students to learn.

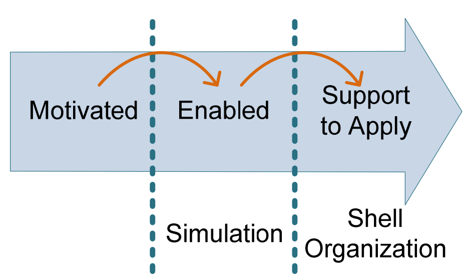

Figure 44 Organizational Learning Support

Interview Rutger Deenen

The following statements have been derived from the interview on Thursday 28 March 2013, in this interview three questions were of particular interest.

On the question if the developed evaluation methodology is applicable for the evaluation of the serious games that are used by Accenture, the following statements were noted:

- Yes, the methodology in general is good and applicable. However, it would be useful to extend the methodology to the evaluation on multiplayer games in a physical environment since most serious games used by Accenture are role-playing games or games were people are interacting in a classroom setting. Such workshops are typically capacitated between 15 to 50 players.
- An important difference between the games Accenture uses and Hazard Recognition is that the serious games Accenture uses are predominantly of a higher level of abstraction. An evaluation method should be able to cope with that.
- The objective of serious games that Accenture uses are to improve behavioral elements of participants such as; cooperation, communication, leadership, teamwork, customer focus, decisiveness. An evaluation method should be capable of discovering these elements in a previous mentioned setting.
- Thus, the workshops mainly train soft skills and it is important to understand how the efficacy of such games can be proven.
- In addition, it would be valuable to have a method available that could understand the efficacy of games that train participants' organizational skills.
- Every month Accenture hosts serious games as part of their new hire days, an introduction for new employees. Above mentioned skills are trained with various games. Accenture is particularly interested in the learning retention of these games a future research could easily be accommodated for such studies for a length of several months.

Appendix IV

During the workshops the players were confronted with several bugs that were present in the game. These bugs have been analyzed by the researcher and are listed below.

Working on heights

1) The player recognizes the workers on the aerial platform then the circle in the checkbox starts spinning meaning that the hazard process is initiated. But the workers are not moving, so there is nothing to see for the player. However, regardless if the player stays or leaves, after a minute or so the movie is initiated that shows the worker falling off the platform. So in the background the simulation starts the hazard procedure, but doesn't visualize it.
2) In this level, sometimes it happens that a hazard is recognized and there is taken action by the player, but the menu gives feedback that there is nothing wrong. Leaving the hazard untreated. However, that hazard does fire.

Lifting and hoisting

1) If the strong winds is not recognized by the player, the hazard fires. However, instead of letting the player start where he left in the game, the game requires to the pre-job discussion again.
2) The hazard of the gap in the barrier tape fires too quick and is sometimes out of sync. Also when this hazard does fire, the movie often shows only a moving crane and no avatar walking and being hit. Meaning that the movie only confuses players.
3) After the pre-job discussion; the player recognizes the gap in the barrier tape shortly after finishing the pre-job discussion and hits space-bar, the simulation freezes.
4) Once the hazard of the crooked crane has fired once, a movie shows the result. If the player recognizes the second time the crooked crane, sometimes it happens that the simulation doesn't recognizes the hazard. Meaning that in the menu a question mark appears instead of a check mark. If that happens, the simulation freezes and it has to be rebooted.
5) There is an inconsistency in the game, the barrier tape is also called marking tape.
6) The load handlers are to be said that they are not communicating properly. However, they should not communicate at all according to procedure. Also, when this is reported they are still signaling.

Appendix V

Student Questionnaire Pre

Date of birth	
Age	
Sex	
Nationality	
Highest level of education	
Educational disciplinary background	
Name University	
Academic background in gaming	
Have you played this simulation before?	
If so, how many times?	

Background Gaming						
How often in your private capacity (ie. not part of education or work) do you play regular games, like board games and role plays?	1. never	2. few times/year	3. monthly	4. Weekly	5. Daily	Don't know
How often in your private capacity (i.e. not part of education or work) do you play computer games, such as pc/internet/mobile games and/or consoles?	1. never	2. few times/year	3. monthly	4. Weekly	5. Daily	Don't know
How often in your study/education/work have you participated in a simulation game (e.g. a management game, business game, policy game, crisis simulation, role play etc., with or without the use of a computer)?	1. never	2. few times/year	3. monthly	4. Weekly	5. Daily	Don't know
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
In general, I think the use of simulation games in education is valuable.	1	2	3	4	5	6
In general, I think it's fun to take part in a simulation game in education.	1	2	3	4	5	6
I think that simulation games in education add something to other forms of teaching (e.g. formal lectures and seminars).	1	2	3	4	5	6

Attitude towards Hazards	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
I would always secure a ladder when using it	1	2	3	4	5	6
I would use fall protection for cutting high trees	1	2	3	4	5	6
I would ask someone to be an extra pair of eyes for cutting complex tree structures	1	2	3	4	5	6
I would use a helmet when hoisting a piano to the second floor's window	1	2	3	4	5	6
I would first thouroghly inspect hoisting equipment before using it	1	2	3	4	5	6
I would use an improvised construction of a table and chair to change a light bulb	1	2	3	4	5	6

Skills	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **detect** hazards	1	2	3	4	5	6
2. I am capable of **detecting** hazards	1	2	3	4	5	6
3. I am able to **detect** hazards	1	2	3	4	5	6
4. I feel able to meet the challenge of **detecting** hazards	1	2	3	4	5	6
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **recognise** hazard patterns	1	2	3	4	5	6
2. I am capable of **recognizing** hazard patterns	1	2	3	4	5	6
3. I am able to **recognise** hazard patterns	1	2	3	4	5	6
4. I feel able to meet the challenge of **recognizing** hazard patterns	1	2	3	4	5	6
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **respond** to hazards	1	2	3	4	5	6
2. I am capable of **responding** to hazards	1	2	3	4	5	6
3. I am able to **respond** to hazards	1	2	3	4	5	6
4. I feel able to meet the challenge of **responding** to hazards	1	2	3	4	5	6

Student Questionnaire Post

Date of birth	
Age	
Sex	

Attitude towards Hazards

	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
I would always secure a ladder when using it	1	2	3	4	5	6
I would use fall protection for cutting high trees	1	2	3	4	5	6
I would ask someone to be an extra pair of eyes for cutting complex tree structures	1	2	3	4	5	6
I would use a helmet when hoisting a piano to the second floor's window	1	2	3	4	5	6
I would first thouroghly inspect hoisting equipment before using it	1	2	3	4	5	6
I would use an improvised construction of a table and chair to change a light bulb	1	2	3	4	5	6

Skills

	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **detect** hazards	1	2	3	4	5	6
2. I am capable of **detecting** hazards	1	2	3	4	5	6
3. I am able to **detect** hazards	1	2	3	4	5	6
4. I feel able to meet the challenge of **detecting** hazards	1	2	3	4	5	6
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **recognise** hazard patterns	1	2	3	4	5	6
2. I am capable of **recognizing** hazard patterns	1	2	3	4	5	6
3. I am able to **recognise** hazard patterns	1	2	3	4	5	6
4. I feel able to meet the challenge of **recognizing** hazard patterns	1	2	3	4	5	6
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **respond** to hazards	1	2	3	4	5	6
2. I am capable of **responding** to hazards	1	2	3	4	5	6
3. I am able to **respond** to hazards	1	2	3	4	5	6
4. I feel able to meet the challenge of **responding** to hazards	1	2	3	4	5	6

Game Design

	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
The instructions and explanations at the start of the simulation game were clear.	1	2	3	4	5	6
The tasks in the simulation game were understandable and clearly described.	1	2	3	4	5	6
The rules of the game were clear and straightforward.	1	2	3	4	5	6
The game materials were understandable and clearly written.	1	2	3	4	5	6

	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
The role we played in the simulation game was understandable and clearly described.	1	2	3	4	5	6
The assignments (tasks) in the simulation game were understandable and clearly described.	1	2	3	4	5	6
All of the materials and documents needed to play the simulation game were available.	1	2	3	4	5	6
Considering the aims of the game, it was sufficiently detailed.	1	2	3	4	5	6
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
Given the aims of the simulation game, the simulation was sufficiently realistic.	1	2	3	4	5	6
The issues in the game represent the challenges in recognition of hazards accordingly.	1	2	3	4	5	6
The simulation game was built up in an interesting and motivating way.	1	2	3	4	5	6

immersion

When playing the game…	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
I lost track of time	1	2	3	4	5	6
Things seem to happen automatically	1	2	3	4	5	6
I felt different	1	2	3	4	5	6
I felt scared	1	2	3	4	5	6
When playing the game…	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
the game feels real	1	2	3	4	5	6
if someone talked to me, I didn't hear them	1	2	3	4	5	6
I got wound up	1	2	3	4	5	6
Time seemed to kind of stand still or stop	1	2	3	4	5	6
When playing the game…	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
I felt spaced out	1	2	3	4	5	6
I didn't answer when someone talks to me	1	2	3	4	5	6
I couldn't tell if I was getting tired	1	2	3	4	5	6
playing seems automatic	1	2	3	4	5	6
When playing the game…	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
my thoughts went fast	1	2	3	4	5	6
I lost track of where I am	1	2	3	4	5	6
I played without thinking how to play	1	2	3	4	5	6
playing made me feel calm	1	2	3	4	5	6
When playing the game…	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
I played longer than I meant to	1	2	3	4	5	6
I really got into the game	1	2	3	4	5	6
I felt like I just can't stop playing	1	2	3	4	5	6

Trade-off	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
The tasks in the simulation game were too easy.	1	2	3	4	5	6
The tasks in the simulation game were too difficult.	1	2	3	4	5	6
The structure of the game was too predefined.	1	2	3	4	5	6
I experienced a balance between difficulty of the tasks and rewards I gained	1	2	3	4	5	6

Learning Fun	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
The game was fun	1	2	3	4	5	6
The exposure to a different industry (chemical sites) was interesting	1	2	3	4	5	6
It was fun to learn things about hazard recognition	1	2	3	4	5	6
I learned much about the recognition of hazards	1	2	3	4	5	6

Bug perception	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
In my perception, the game was bug free	1	2	3	4	5	6
The bugs I noticed were of negative influence on my performance as Supervisor	1	2	3	4	5	6
The bugs I noticed took me out my game flow	1	2	3	4	5	6
Yes I noticed bugs, but they didn't influenced my gameplay or fun	1	2	3	4	5	6

Knowledge	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **explain** in detail about hazard recognition	1	2	3	4	5	6
2. I am capable of **explaining** and sharing knowledge in detail on hazard recognition	1	2	3	4	5	6
3. I am able to **explain** into depth about the hazards on work sites	1	2	3	4	5	6
4. I feel able to meet the challenge of mastering knowledge on hazards	1	2	3	4	5	6
5. I feel that I have extended knowledge on hazards	1	2	3	4	5	6

6. Name maximum 10 types of typical hazards on worksites	1
	2
	3
	4
	5
	6
	7
	8
	9
	10

During a hoisting & lifting procedure…	Multiple answers may be correct
The banksman should have eye contact with:	A) The crane driver B) Lifting personall C) Supervisor D) No one
Who is responsible for signaling the cranedriver	A) Banksman B) Supervisor C) Lifting personall D) No one
With heavy winds, you can work on a	A) Cherry Picker B) Ladder C) Crane D) Not on Heights
Lifting straps should be renewed every	A) 6 months B) 12 months C) 18 months D) 24 months
The banksman is allowed to	A) Give hoisting directions B) Smoke C) Help with the load D) Ask for permits

Professional Questionnaire Pre

Date of birth	
Age	
Sex	
Nationality	
Country of profession	
Highest level of education	
Educational disciplinary background	
Years of professional experience	
Number of years within Oil/Gas Industry	
Have you played this simulation before?	
If so, how many times?	

Background Gaming						
How often in your private capacity (ie. not part of education or work) do you play regular games, like board games and role plays?	1. never	2. few times/year	3. monthly	4. Weekly	5. Daily	Don't know
How often in your private capacity (i.e. not part of education or work) do you play computer games, such as pc/internet/mobile games and/or consoles?	1. never	2. few times/year	3. monthly	4. Weekly	5. Daily	Don't know
How often in your study/education/work have you participated in a simulation game (e.g. a management game, business game, policy game, crisis simulation, role play etc., with or without the use of a computer)?	1. never	2. few times/year	3. monthly	4. Weekly	5. Daily	Don't know

Professional Background	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
In my daily work, I'm actively involved in hazardous situations?	1	2	3	4	5	6
In my daily work, I'm actively involved in hazard detection?	1	2	3	4	5	6
In my daily work, I'm actively involved in responding to hazards?	1	2	3	4	5	6
Before I took part in this simulation game, I was already experienced and informed about issues related to hazards	1	2	3	4	5	6
I'm a subject expert in my field of profession	1	2	3	4	5	6
Hazards are part of my every day working environment	1	2	3	4	5	6

Attitude towards Hazards	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
I would always secure a ladder when using it	1	2	3	4	5	6
I would use fall protection for cutting high trees	1	2	3	4	5	6
I would ask someone to be an extra pair of eyes for cutting complex tree structures	1	2	3	4	5	6
I would use a helmet when hoisting a piano to the second floor's window	1	2	3	4	5	6
I would first thouroghly inspect hoisting equipment before using it	1	2	3	4	5	6

I would use an improvised construction of a table and chair to change a light bulb	1	2	3	4	5	6

Skills	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **detect** hazards	1	2	3	4	5	6
2. I am capable of **detecting** hazards	1	2	3	4	5	6
3. I am able to **detect** hazards	1	2	3	4	5	6
4. I feel able to meet the challenge of **detecting** hazards	1	2	3	4	5	6
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **recognise** hazard patterns	1	2	3	4	5	6
2. I am capable of **recognizing** hazard patterns	1	2	3	4	5	6
3. I am able to **recognise** hazard patterns	1	2	3	4	5	6
4. I feel able to meet the challenge of **recognizing** hazard patterns	1	2	3	4	5	6
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **respond** to hazards	1	2	3	4	5	6
2. I am capable of **responding** to hazards	1	2	3	4	5	6
3. I am able to **respond** to hazards	1	2	3	4	5	6
4. I feel able to meet the challenge of **responding** to hazards	1	2	3	4	5	6
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to stay in **control** in hazardous environments	1	2	3	4	5	6
2. I am capable of staying in **control** in hazardous environments	1	2	3	4	5	6
3. I am able to stay in **control** in hazardous environments	1	2	3	4	5	6
4. I feel able to meet the challenge of being in **control** of hazardous environments	1	2	3	4	5	6
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to have **oversight** in hazardous environments	1	2	3	4	5	6
2. I am capable of having **oversight** in hazardous environments	1	2	3	4	5	6
3. I am able to having **oversight** in hazardous environments	1	2	3	4	5	6
4. I feel able to meet the challenge of having **oversight** of hazardous environments	1	2	3	4	5	6

Knowledge	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **explain** in detail about hazard recognition	1	2	3	4	5	6
2. I am capable of **explaining** and sharing knowledge in detail on hazard recognition	1	2	3	4	5	6
3. I am able to **explain** into depth about the hazards on work sites	1	2	3	4	5	6

4. I feel able to meet the challenge of mastering knowledge on hazards	1	2	3	4	5	6
5. I feel that I have extended knowledge on hazards	1	2	3	4	5	6
6. Name maximum 10 types of typical hazards on worksites	1					
	2					
	3					
	4					
	5					
	6					
	7					
	8					
	9					
	10					

During a hoisting & lifting procedure…	
The banksman should have eye contact with:	A) The crane driver B) Lifting personall C) Supervisor D) No one
Who is responsible for signaling the cranedriver	A) Banksman B) Supervisor C) Lifting personall D) No one
With heavy winds, you can work on a	A) Cherry Picker B) Ladder C) Crane D) Not on Heights
Lifting straps should be renewed every	A) 6 months B) 12 months C) 18 months D) 24 months
The banksman is allowed to	A) Give hoisting directions B) Smoke C) Help with the load D) Ask for permits

Professional Questionnaire Post

Date of birth	
Age	
Sex	

Attitude towards Hazards	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
I would always secure a ladder when using it	1	2	3	4	5	6
I would use fall protection for cutting high trees	1	2	3	4	5	6
I would ask someone to be an extra pair of eyes for cutting complex tree structures	1	2	3	4	5	6
I would use a helmet when hoisting a piano to the second floor's window	1	2	3	4	5	6
I would first thouroghly inspect hoisting equipment before using it	1	2	3	4	5	6
I would use an improvised construction of a table and chair to change a light bulb	1	2	3	4	5	6

Skills	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **detect** hazards	1	2	3	4	5	6
2. I am capable of **detecting** hazards	1	2	3	4	5	6
3. I am able to **detect** hazards	1	2	3	4	5	6
4. I feel able to meet the challenge of **detecting** hazards	1	2	3	4	5	6
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **recognise** hazard patterns	1	2	3	4	5	6
2. I am capable of **recognizing** hazard patterns	1	2	3	4	5	6
3. I am able to **recognise** hazard patterns	1	2	3	4	5	6
4. I feel able to meet the challenge of **recognizing** hazard patterns	1	2	3	4	5	6
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **respond** to hazards	1	2	3	4	5	6
2. I am capable of **responding** to hazards	1	2	3	4	5	6
3. I am able to **respond** to hazards	1	2	3	4	5	6
4. I feel able to meet the challenge of **responding** to hazards	1	2	3	4	5	6
	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to stay in **control** in hazardous environments	1	2	3	4	5	6
2. I am capable of staying in **control** in hazardous environments	1	2	3	4	5	6
3. I am able to stay in **control** in hazardous environments	1	2	3	4	5	6
4. I feel able to meet the challenge of being in **control** of hazardous environments	1	2	3	4	5	6

	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable

1. I feel confident in my ability to have **oversight** in hazardous environments	1	2	3	4	5	6
2. I am capable of having **oversight** in hazardous environments	1	2	3	4	5	6
3. I am able to having **oversight** in hazardous environments	1	2	3	4	5	6
4. I feel able to meet the challenge of having **oversight** of hazardous environments	1	2	3	4	5	6

Knowledge	Strongly disagree	disagree	neutral	agree	strongly agree	Not applicable
1. I feel confident in my ability to **explain** in detail about hazard recognition	1	2	3	4	5	6
2. I am capable of **explaining** and sharing knowledge in detail on hazard recognition	1	2	3	4	5	6
3. I am able to **explain** into depth about the hazards on work sites	1	2	3	4	5	6
4. I feel able to meet the challenge of mastering knowledge on hazards	1	2	3	4	5	6
5. I feel that I have extended knowledge on hazards	1	2	3	4	5	6
6. Name maximum 10 types of typical hazards on worksites	1					
	2					
	3					
	4					
	5					
	6					
	7					
	8					
	9					
	10					

During a hoisting & lifting procedure…	
The banksman should have eye contact with:	A) The crane driver B) Lifting personall C) Supervisor D) No one
Who is responsible for signaling the cranedriver	A) Banksman B) Supervisor C) Lifting personall D) No one
With heavy winds, you can work on a	A) Cherry Picker B) Ladder C) Crane D) Not on Heights

Lifting straps should be renewed every	A) 6 months B) 12 months C) 18 months D) 24 months
The banksman is allowed to	A) Give hoisting directions B) Smoke C) Help with the load D) Ask for permits

"Thank you so much for playing my game!"

(Super Mario, June 26, 1996)