Employing Variation in the Object of Learning for the Design-based Development of Serious Games that Support Learning of Conditional Knowledge

Martin Petkov Ruskov

University College London

for a PhD Degree in Computer Science

I, Martin Petkov Ruskov confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

The signed declaration should be followed by an abstract consisting of no more than 300 words. An extra copy of the abstract typed on the special form provided with the entry form is required for publication in the Index to Theses. The should be submitted with the final copy of your theses.

Abstract

Learning how to cope with tasks that do not have optimal solutions is a life-long challenge. In particular when such education and training needs to be scalable, technologies are needed to support teachers and facilitators in providing the feedback and discussion necessary for quality learning. In this thesis, I conduct *design-based research* by following a typical game development cycle to develop a *serious game*. I propose a framework that derives learning and motivational principles to include them into the design of *serious games*. My exploration starts with *project management* as a learning domain, and for practical reasons, shifts towards *information security*.

The first (concept) phase of the development includes an in-depth study: a *simulation game* of negotiation (Study 1: class study, n=60). In the second (design) phase I used *rapid prototyping* to develop a gamified web toolkit, embodying the CCO framework from *crime prevention*, making five small-scale formative evaluations (Study 2, n=17) and a final lab evaluation (Study 3, n=28). In the final (production) stage the toolkit was used in two class studies (Study 4, n=34 and Study 5, n=20), exploring its adoption in a real-world environment.

This thesis makes three main contributions. One contribution is the adaptation of the iterative method of the phenomenographic *learning study* to the study of the efficiency of *serious games*. This employs open questionsing, analysed with 3 different means of analysis to demonstrate 4 distinct types of evidence of *deep learning*. Another contribution is the provided partial evidence for the positive effects from the introduction of *variation* on engagement and learning. The third contribution is the development of four *design-based research* principles: i) the importance of being agile; ii) feedback from interpretation of the theory; iii) particular needs for facilitation; and iv) reusing user-generated content.

Acknowledgements

First and foremost, I would like to thank the people who have helped me shape and complete this research project of mine – my supervisors **Prof. M. Angela Sasse** and **Prof. Paul Ekblom**. Working with you was truly inspiring.

In various periods of my work I have taken ideas, direction and criticism from five other colleagues of mine – **Prof. Mick Flanagan**, **Dr Charlene Jennett**, **Dr Paul Walker**, **Dr Manuel Fradinho Oliveira** and **Dr Will Seager**. There is a major part of my research that was strongly influenced by ideas I have discussed with each of you.

A number of student peers helped me with different smaller bits, mostly **Dr Miguel Malheiros**, **Jose Maestre Celdran** and **Zhanelya Subebayeva**, but also **Steve Dodier-Lazaro**, **Horace Li**, **Maryum Styles**, **Daniyar Kaliev** and **Mohammed Zaman**.

Nina, **Deni** and **Boyana** were very helpful with their proofreading and sanity-checks. Thank you for the time to work on something so different from what you do yourselves.

And on a personal note, I want to thank **my parents and all my other teachers and mentors** that have served as my role models and have encouraged me throughout my entire life. I would not have reached this point if I did not have your guidance along my path.

Table of Contents

Index of Figures

Index of Tables

Abbreviations

ACTA – Applied Cognitive Task Analysis
ANOVA – Analysis of Variance
CAF – Conceptual Assessment Framework
CALT – Centre for Advancement of Learning and Teaching
CCO – Conjunction of Criminal Opportunity
CISO – Chief Information Security Officer
CoP – Communities of Practice
CPA – Competence Performance Analyser
CS – Computer Science
CSCL – Computer-Supported Collaborative Learning
DPE – Design, Play, and Experience
ECD – Evidence-Centred Design
GBL – Game-Based Learning
HCI – Human-Computer Interaction
IDS – Interactive and Digital Storytelling
iGEQ – in-Game Engagement Questionnaire
IL – Inquiry Learning
IS – Information Security
ITS – Intelligent Tutoring Systems
MCQ – Multiple-Choice Questions
MMOG – Massively-Multiplayer Online Games
MMORPG – Massively-Multiplayer Online Role-Playing Games
MSI – Management Science and Innovation
NLP – Natural Language Processing
NPC – Non-Player Character
OKEI – Organisation, Knowledge, Environment, Individual
PBL – Problem-Based Learning
PM – Project Management
PMI – Project Management Institute
RPG – Role-Playing Games
SAL – Self-Assessed Learning
SME – Subject-Matter Expert
SOLO – Structure of Observed Learning Outcomes
TARGET – Transformative, Adaptive, Responsive and enGaging EnvironmenT
TEL – Technology-Enhanced Learning

Chapter 1: Introduction

1.1 Current Challenges

In our highly networked and knowledge-based society, the increasing speed of change is a steady source of new challenges. People and their individual capabilities are a key driving force in addressing emerging challenges. In a society where one size does not fit all, there is a need to grow beyond the habits of faster-better-cheaper grow into life-long learning and reflective practice. The increasing technological and interactional complexity requires critical learning and expertise development in response. *Technology-enhanced learning*[1] already provides a number of opportunities for practice-based and networked learning. Computer games in particular, have shown to be very good at motivating their target audience (Ryan et al., 2006) and helping people develop deeper understanding of complex phenomena (Hays, 2005). However, not all educational computer games (commonly referred to as *serious games*) are effective (even less so efficient) in supporting motivation to learn and learning itself. In this thesis I endeavour in an exploration to understand and improve both engagement and learning with serious games.

Research in three particular domains can contribute towards these goals. These are contributions from research in games and simulations; forms of complex knowledge and variation theory of learning. Recent contributions in these three areas are introduced in the next section and revisited in greater detail in Chapter 2.

1.2 Serious Games and Complex Knowledge

Within computer science the discipline of serious games research falls in the intersections of the wider disciplines of *human-computer interaction* (HCI) and *technology-enhanced learning*, as studies on a particular form of visual interactive media used to support learning (see Figure 1). Typically for the wider discipline, studies on serious games focus on how they could be developed (a case of *design-based research*) and how does their use affect users (usability and learning evaluation studies). The work reported in this thesis contributes to scientific knowledge in both building serious games, using design-based research and evaluating their effects for engagement and learning in particular.

Serious games are instances of virtual environments. As such, they can provide a controlled setting for experiencing and learning tasks without the distractions,

1 Here, and throughout this thesis, terminology from different research domains is used in italics. Explanation of these terms and abbreviations can be found in the Glossary at the end of the thesis.

complexities and risks of the real world. Their interactivity, in line with principles of *constructivism* (Hmelo-Silver et al., 2007), allows for a much more engaging environment than the more predetermined and linear traditional lectures and presentations. Games and simulations are experiential in nature. Research on self-emerged communities comprised of players of massive-multiplayer online games (Reeves and Malone, 2007), has shown that although online games employ fictitious characters and stories, people playing develop complex capabilities that are also relevant in the real world. When considering learning with games is a case of *problem-based learning* (PBL), this could be related to the findings reported by Gijbels and colleagues (2005). These researchers concluded that PBL tends to be less efficient when learning *declarative knowledge* (*know-what*), but outperforms traditional learning techniques when considering *procedural* (*know-how*) and *conditional knowledge* (called *know-with* by Broudy (1977) according to (Bransford and Schwartz, 1999)). Although Gijbels's findings on PBL are indicative, it is questionable to what extent they could be considered transferable to serious games for learning. A detailed discussion of different forms of learning and knowledge is presented in Section 2.1.

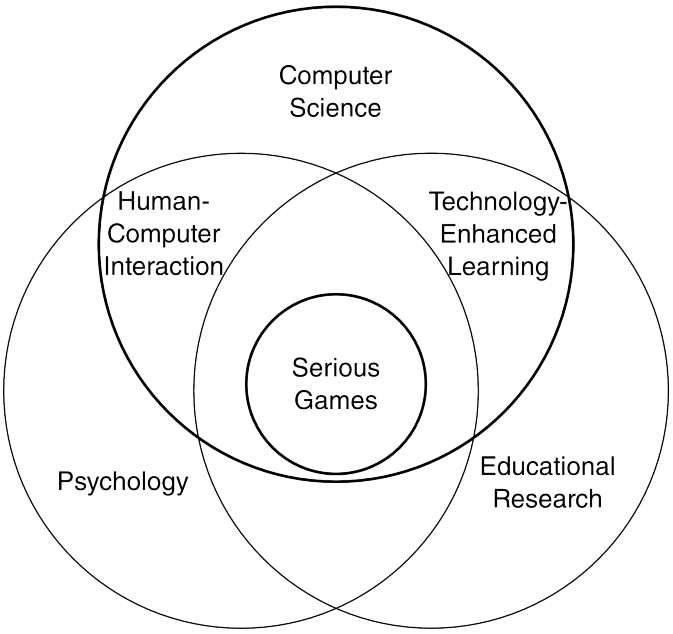


Figure 1: Venn diagram of Serious Games and other related disciplines

Whenever interpersonal interactions are involved, it can be generally expected that declarative and procedural knowledge are capturing only a portion of the puzzle. When dealing with other people, a type of conditional knowledge is critical to successful problem-solving. This is most apparent when *soft-skills* are being addressed. Interpersonal interactions and soft-skills are increasingly attracting attention in a range of management and governance professions, notably project management and information security, which are the two subject domains I will focus on in this thesis. In project management an example of similar troublesome knowledge is the distinction between interest and positions of negotiating parties (Fisher and Ury, 1991; Flanagan et al., 2010). The game

vLeader studied in Chapter 5 makes explicit the distinction of interest and position in negotiations. In information security a challenge to thinking, is the need for more complex view than easily perceived (Ekblom, 2011a). Adding these further interpersonal aspects to security, leads not only to more factors to consider, but a conditional mode of thinking about their interplay. The CCO framework is an attempt to capture such complexities and guide learners to solve them (Ekblom, 2011a). The development process is reported in Chapter 6. Evaluations were conducted in Chapter 7 and Chapter 8.

The *autonomy* provided by serious games seems to be one of their greatest advantages. Delegating choices to players is the core technique that games use to achieve their inherent interactivity and non-linearity. In order for this to happen, games need to be able to support a certain degree of variation that players can experience. This variation may be expressed in different ways in the broad universe of games, but its importance in the field has not been left unnoticed. Jesper Juul (2007), a prominent game designer and theorist, has developed an extended hypothesis for the attractive value of variation. He draws a connection between variation over time in games and the maintained attractiveness of these games (Juul, 2007). He elaborates on the attractiveness of gradually increasing variation with the progression of a game. If Juul's claims prove to be valid over the whole domain of games, they would turn into a very strong argument towards the use of variation in learning with games.

1.3 Learning from Variation in Serious Games

Coming from the context of education, a number of different theories (Marton and Pang, 2006; Meyer et al., 2008; van Merriënboer and Sweller, 2005) have argued for the importance of variation in learning. Research in this area has accumulated evidence that learners understand the *object of learning* (which is the phenomenon they explore) better when they experience it in various ways, in other words, get exposed to its variations. Often, in order to demonstrate variation to learners, educators and researchers have used the support of interactive tools (Pang and Marton, 2003), simulations (Fraser and Linder, 2009) or even traditional games (Schilling et al., 2003). That is what prompted me to try to explore what are possible systematic ways to introduce learning variation in games, and what significant learning effects can be observed and identified.

Considering the current situation in research and development of serious games, work in two main directions is necessary. The first is investigation of cognitive learning with serious games to establish which forms of learning are effective and efficient tools to

support engagement and learning. The second is providing the development to support such engagement and learning, both in terms of development and evaluation methodologies for serious games. Reviews of current empirical research, not only in simulations (Anderson and Lawton, 2009) and serious games (Hays, 2005), but also in other related forms of experiential learning, like PBL (Gijbels et al., 2005), thus showing partial evidence of when these are useful. Notably, there seems to be differing degrees of success for different learning and assessment techniques used. In the cited meta-reviews there is little empirical evidence that experiential learning techniques are appropriate for more complex learning, depending on the user context. In the case of the review by Gijbels and colleagues (2005), PBL appears to be more successful for procedural and conditional knowledge, whereas Anderson and Lawton (2009) associate business simulations with higher levels of understanding according to Bloom's cognitive domain taxonomy. In Section 2.1 knowledge and levels of understanding are discussed in more detail.

Another common finding across reviews is the differing levels of success that different games have. In a review Kebritchi and Hirumi (2008) noted that very often *educational games* do not live up to the promises of the theoretical support they claim. Although designers often claim that a certain game is rooted in a particular pedagogical theory, its authors generally fail to show how they instantiated the theory. On the other hand, researchers in educational design, Davies and Mangan (2006) suggested that the design of learning activities should undergo a process from theory through to pedagogical principles in order to finally be able to arrive at the actual learning activities. Currently, such a systematic game design approach does not exist, and thus is an open research challenge.

1.4 Research Goals

The goal of the current research project is to investigate opportunities to improve learning of conditional knowledge, and interpersonal skills, with the help of serious games. I intend to achieve this by studying learning variation in an existing serious game, and ultimately, systematically introducing it into the design of a new one, as developed here. As explained in further detail in the subsequent section, I will use design-based research as a scientific reflective practice for this development.

In line with established practice in serious games, variation should contribute to learning – various literature sources have already indicated both effects in the introduction of variation. Furthermore, there are indications that variation as a principle, which on top of its widely acknowledged pedagogical value, could also contribute to learner motivation.

Part of the hereby declared goal is to get a deeper understanding of learning activities that converge around variation in game-based learning, including their motivational value. In particular, the goal is to examine with rigour and detail on how the application of variation theory can improve effectiveness in developing skills, relevant for the work environment, and what are reliable and valid ways to evaluate such developments. This research project considers graduate students as participants, as they are assumed to be an adequate proxy for young professionals.

In order to achieve these goals, three research questions are being explored in this thesis:

- RQ1: How does experiencing variation in a serious game contribute to improved understanding of learning objectives?
- RQ2: How does experiencing variation in a serious game contribute to improved application of learning objectives?
- RQ3: How does experiencing variation in a serious game contribute to higher levels of engagement with the serious game?

As a way to address these research questions, here I explored variations of different manifestations of the *object of learning*. Three specific manifestations of the object of learning are elaborated within *variation theory* (Pang and Marton, 2003). The first one is the *intended object of learning* is the preparation and design of the learning activity. In typical educational research studies this includes learning materials and any preparations that teachers make while planning the learning activity. The second manifestation – the *enacted object of learning* – is the actual way the learner gets to experience what is learned. Typically in class studies this is class interaction, for example as it could be recorded on camera. The third and final manifestation – the *lived object of learning* – is the experience as the learners "see, understand and make sense" (Marton et al., 2004) of what is learned. In a class environment a discussion will generally take place to try and closely align the lived object of learning and the intended object of learning.

In a development very similar to variation theory's manifestations of the object of learning, but in the domain of serious games, Brian Winn (2008) puts forward the Design-Play-Experience (DPE) framework, where the design corresponds to the intended object of learning; play to the enacted object of learning; and experience to the lived object of learning. The DPE framework provides a validation of the approach of embedding the object of learning in a serious game taken here. Research on the object of learning and the

DPE framework are discussed in Chapter 2, and my approach is explained in greater detail in Section 4.1.1.

1.5 Research Approach

I examined the effects that variation has on the different objects of learning on one side, and on understanding, application and engagement on the other. This provided multifaceted insights towards the three research questions: effects on learning (RQ1), application (RQ2) and engagement (RQ3). As a way to enable learning in context (which is *situated learning*), this work aims at the development of a simplified serious game that would provide an environment for virtual practice, where learners are able to develop skills that should be transferable to solving real-world problems. In other words, they would know when, where, how and who to involve. The three research questions, as formulated above, still do not contain enough specificity to predetermine the methods used for each of them. To clarify this in short, here it is briefly explained how each of these were approached in this thesis. This is further discussed in Chapter 4.

Considering the practical constraints of PhD research, the work reported in this thesis follows a development lifecycle: pilot evaluation with of pre-existing game, *design-based research* (also referred to as *development research* because of emphasis on reflective development) and real-world evaluations of the developed prototype (see Table 1 with list of studies).

In a reflection on pedagogical approaches, consulted with the reviewed literature and summarised in Section 4.1.1, for the evaluations I adapted my method to explore forms of

Phase	Study (Chapter)	Measured Construct	Type	n	Measures/ Analysis
Concept	**1 (5)**	learning and engagement (negotiation)	class study with a proxy	60*	open questions, MCQ, iGEQ, role-play
Design	**2 (6)**	usability	usability study (formative)	17	usability heuristics
	3 (7)	learning and engagement	lab study (summative)	28	open questions, MCQ, iGEQ
Production	**4 (8)**	engagement	class study	34*	engagement survey
	5 (8)	engagement	class study	25*	engagement survey

Table 1: Studies conducted within this thesis and key information about them. Here, and throughout this thesis, N indicates the number of participants.

* the indicated studies were conducted in a class and not all students provided all types of collected data. Further details on the participating number in each measure can be found in the corresponding chapters.

deep learning from the *phenomenographic learning study* (explained in Section 4.1.1). Consequently, I utilised an approach influenced by *design-based research* (Anderson and Shattuck, 2012) and developed a prototype for a serious game of my own, called *CCO toolkit*. This approach is much less conclusive than experimental studies, but it better addressed the difficulties associated with the study of *technology-enhanced learning,* which is an intersection of educational research and computer science in a real-world setting.

I conduct my research with graduate students, a group representative of young professionals, who are familiar with the practice of life-long learning. The first study reported in this thesis utilised an existing game called *vLeader*, in a class of graduate students in a university module on leadership. The study aimed to pilot the evaluation methodology which was later used to evaluate the *CCO toolkit*. Yet, it also underlined the necessity to have access to the programming code in order to modify the software. Due to the limited scope of this thesis, software was developed only to the level of a minimalistic prototype. In this prototype a number of features were implemented only superficially, such as method stubs or mock objects, which are software interfaces that delivered what was required, but a minimally feasible way. Beyond development of the prototype, work on CCO included evaluations – one in lab settings and two class evaluations with graduate students on a module on human factors in information security.

Based on the expectation that serious games will be more efficient when addressing complex forms of knowledge, and the related need to engage learners beyond surface learning, my measures of learning needed to focus on deep learning. To ensure this, I used open written questioning as a multi-layered measure of learning. Open-answer questions are a highly insightful assessment method (Atherton, 2013a; Pang and Marton, 2007) and is widely used in all levels of institutional education. Due to the limited scope of a PhD thesis that also includes *development research*, I did not go as far as validating the assessment by a second coder. This was complemented by multiple-choice questions (MCQ), because these are the established and widely accepted ways to measure learning (Nicol, 2007; Scouller, 1998). However, the exploratory and often class-based nature of these studies, leading to comparably small sample sizes, did not make for high expectations of being able to test for strong statistical significance in the results.

In terms of real-world benefits actual performance is more important than demonstration of knowledge. This is why in the pilot study (Study 1 conducted in a class using the game

Research Goal	Findings
1. Study how learning of conditional knowledge can be supported by the introduction of variation in serious games	Development processes: - establishment of a framework leading from theory to game-based learning activities - evidence of different ways in which variation affects learning with serious games
2. Develop a serious game employing a design-based approach and refining design principles for serious games	Design principles: - the importance of being agile - feedback from implementation of the theory - particular needs for facilitation - reusing user-generated content
3. Develop assessment measures that capture deep learning and conditional knowledge	- adaptation of the phenomenographic learning study - method for open questions assessment; to measure deep learning as part of a portfolio of learning and engagement assessment tools

Table 2: Contributions of this thesis. These can be viewed from two perspectives: substantive and methodological, and contributing to research and practice

vLeader) students were also assessed through role-playing simulation scenarios. However, the results of this assessment appeared to be radically different from results in the game. In effect, this did not allow me to draw any conclusions with confidence and led to the abandonment of role-play assessment in subsequent studies. In Study 3 (lab study using the *CCO toolkit*) application was measured by asking learners to list their set of solutions to the problem during the tests – before and after the learning activity. Non-intrusive in-game performance data was also collected and examined for correlation with other established measures, such as interviews, recorded observations and think-aloud sessions. However, validation whether game scoring mechanisms actually represented the learners' ability to apply their knowledge in the subject domain remained beyond the scope of this thesis.

Finally, engagement was also measured in several different ways. Engagement over longer periods was measured by the time and frequency learners used the software. Also, in Studies 1 and 3 a standardised measure (iGEQ) was utilised to gauge short term engagement. Finally, in Study 3, the quantity of generated ideas was used to indirectly represent how much learners were engaged with the toolkit.

1.6 Outline of Contributions

In this thesis, I report my research contributions to both the development and evaluation of serious games which contribute in three distinctive ways. As illustrated in Table 2, I make advancements in knowledge in three directions: i) evidence and analysis of

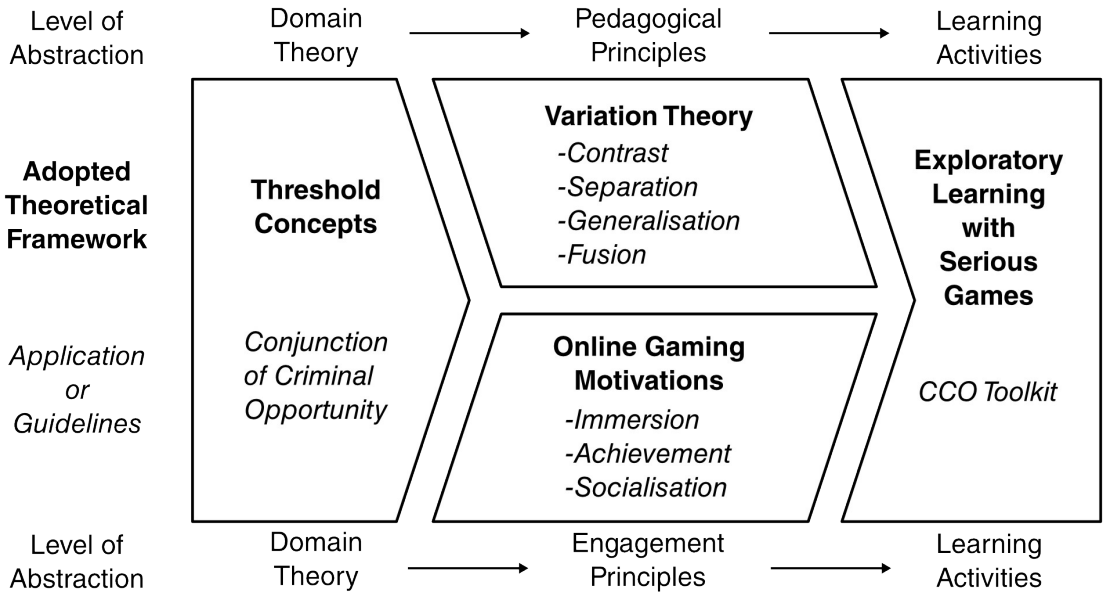


Figure 2: The process of deriving serious games from the theory intended to be taught. This is based on Davies and Mangan's (2006) approach towards threshold concepts, but reflects the fact that serious games need to be, not only beneficial for learning, but also engaging. Regular texts in the figure reflect the framework. In bold are the theoretical frameworks used in this thesis, and in italics are the particular instantiation of these frameworks.

contributions of variation to learning conditional knowledge with serious games; ii) a detailed method for development of serious games addressing conditional knowledge; and iii) a method for the evaluation of deep learning, based on a portfolio of measures, including open written assessment. The contribution related to variation are summarised here. Both the design principles (relevant to the development) and the employed evaluation method are a refinement of the corresponding principles and evaluation methods reviewed from literature (Chapter 2) and derived in the adopted methodology (Chapter 4).

Two key contributions are made to the process of development (practical method). The first of these is a framework for the development of serious games for learning. The second is an adaptation of the game development lifecycle for the purposes of design-based research.

The development framework expands on Davies and Mangan's work (2006) on deriving learning activities from theory intended to be taught. Since serious games have to be both engaging and educational, I expanded the process to also derive the engaging elements of the serious game from the theory, as shown in Figure 2. To actually put this method into practice, I consider *i)* the *Conjunction of Criminal Opportuity* framework as the theory to be learned; and *ii)* apply *variation theory* and a findings about *motivations to play;* iii) to finally establish particular features of the serious game being developed.

The *CCO toolkit* is named after the *Conjunction of Criminal Opportunity* framework (Ekblom 2011), a model that represents criminal situations in much greater richness than the commonly used *problem analysis triangle* of crime prevention (Knutsson and Clarke, 2006), but because of this richness it is more challenging for it to be learned and applied. Because of this specific reason, I consider this to be a potential threshold concept, and a very good candidate to be explored with *serious games*.

Motivation to play games has been widely studied by both practitioners and researchers. By far, the most rigorously constructed typology, expanding on previous theories was developed by Nick Yee (2005), and focuses on online multiplayer games – a genre that closely corresponds to the goal of the practical aspect of this thesis.

The research effort reported in this thesis provides a case study for the implementation of variation in the design of serious games. This research effort attempted to utilise the four conditions for learning, identified by Marton and Tsui (2004): *contrast*, *separation*, *generalisation*, *fusion* and described in greater detail in Subsection 2.1.1. They were used as a channel to deliver variation the concepts of interest for learners to experience.

The game development lifecycle consists of three phases listed in the first column of Table 1. As the first step of the current research, exploratory studies with existing simulation games have been conducted to establish the intended concept. Subsequently, two attempts were made to develop a pilot, as a dedicated game to address previously identified learning goals. The practical scope of this research is the design and development of the resulting serious game. This was the *CCO toolkit* – a low-fidelity prototype of a serious game that is intended to be used in class in conjunction with other materials and facilitator support. This use was piloted in the final production phase of development. As a consequence to factors external to this research, the learning domain of the pilot study was project management (Section 5.2). The learning domain for the developed toolkit and the studies conducted with it was information security.

In the studies conducted as part of this research thesis, I explored variation in different perspectives on the object of learning as described below. I explored design decisions that represent each of the three manifestations of the object of learning.

In the context of serious games, the *intended object of learning* includes the used game itself, and its learning-related content. The first of these – the game implementation – is also the natural first candidate to have variation introduced by exhibiting its different

aspects. In studies with existing games I had no control over the learning content in the game. This is the reason why in Study 1 (reported in Chapter 5) I introduced variation in paper-based supplementary materials, but many of the participants did not use them. In Studies 3 to 5 (reported in Chapter 7 and 8) variation in the content was introduced by exposing learners to ideas that were new to them. The sample sizes used did not allow statistical significance testing. Despite that there were observably higher number of contributions from lab study participants, who were exposed to variation of the intended object of learning. This was an indicator of better application and engagement. Regarding understanding, some participants reflected on ways that variation had helped them critically reflect on ideas. Being a theory that involves some sort of categorisation (see Section 6.1 for detailed description), the CCO theoretical framework inherently features variation. Participants in the study rationalised how this allowed them to derive, by themselves, properties of the learned framework that were not discussed.

In a social game the *enacted object of learning* includes the interactions that take place within the game, including discussions and comments. Another way to provide a rich and varied enacted object of learning is through engaging learners in role-play. These are the two formulations of the enacted object of learning which I had explored.

In Study 1 (Chapter 5), and Studies 4 and 5 (Chapter 8) class studies students did not engage in any form of conversation beyond the short class sessions, therefore there was no substantial discussion. In Study 3 (Chapter 7) discussion was inherently linked to the introduction of variation in the learning content (which is the *intended object of learning*). Learners were provided with an interface to comment on ideas that they had seen. Whenever learners could comment on ideas contributed by others (the introduced variation) this was an example of discussion. In contrast, whenever they commented on their own ideas (control condition) this was reflection. During the interviews some participants suggested that they found the discussion both interesting and insightful.

In literature, for example in (Wills et al., 2010), role-play is identified as very important for learning. It was essential for all research activities reported in this thesis. However, due to the practical constraints of the software, its introduction took different forms in different studies. In Study 1 it was expressed only as learners assuming the role of the game's protagonist. The length of engagement allowed the developers of *vLeader* to put learners in the shoes of both the employee and the manager in a company, while letting them develop the same character troughout the consecutive levels. For the *CCO toolkit* it was

acknowledged that among the biggest benefits of role-play is that learners get a chance to interpret the learning experience from multiple perspectives. Thus, in the resulting toolkit, learners were led to assume three different roles. This provided an opportunity for study participants to discuss how their learning and engagement improved as a result of role-play.

Obviously, the *lived object of learning* cannot be manipulated directly, but the learner's predisposition and motivation have an impact on the learning experience. As a result, I explored ways to improve them with serious games. These two factors which influence the lived object of learning are also explored in this thesis.

One feature available in *vLeader* allowed learners to choose between playing to exercise themselves or trying to achieve their best performance (as measured by in-game scores). This, along with the fact that the performance score awarded by an established game is inherently a measure of how successful the learners were, allowed me to compare their performance when playing the game in the two different modes. From that I concluded that learners did achieve better scores in performance mode. This could be partially explained by an exploratory study beyond the scope of this thesis that we conducted together with Malheiros (2011), where participants expressed common views on how sharing gaming experience with others impacted their readiness to engage with the game.

The collected user-generated content in the *CCO toolkit* provided an insight into learners' thinking and ideas about the problem which can be valuable to security experts. It also allows for reflection on the theoretical framework behind the toolkit. Such artefacts were collected in this study, but their analyses are beyond the scope of this thesis.

This thesis continues with three review chapters – Chapter 2 explores the literature on learning, games and motivation; Chapter 3 reviews different methodologies previously used to explore questions about learning and engagement and describes the adapted methodology used herein; Chapter 4 presents an overview of state of the art games and simulations in project management and information security. The contribution reported in this thesis is a *design-based research* project. As already outlined in Table 1 this report is separated by the three phases: concept, design and production. The concept phase, reported in Chapter 5, was an evaluation with an established game used as a proxy for the evaluation of the *CCO toolkit* before it was developed. This study was published (Ruskov & Seager 2011).

Supported by my research supervisors, I developed a gamified web toolkit to explore case studies in *information security*. Chapter 6 outlines the rapid-prototyping development process. Chapter 7 reports on a lab evaluation of the first usable prototype of the toolkit, and Chapter 8 is a report of two class evaluations that followed as a next step in the development of the *CCO toolkit*. The development part was published in (Ruskov et al., 2012), evaluation of learning in the lab study was published in (Ruskov et al., 2013) and a publication is pending to report the perceptions in the lab study.

Chapter 9 discusses methods for assessment of deep learning grounded in evidence from the conducted learning assessments. The concluding tenth chapter summarises the contributions within this research, followed by a discussion and avenues for future research.

Chapter 2: Literature Review

This literature review is focused on state of the art substantive research in three directly relevant topics: learning, games and motivation. The first one – learning is considered here as the overall goal that needs to be achieved. It is being explored from the perspective of cognitive load, knowledge classification frameworks for the purposes of assessment, constructivist approaches to learning and other supporting factors. The second section considers research in the context of simulation games as the tool that enables learning. It considers how simulation games fit in other related paradigms and relevant design techniques that could contribute to learning. The third and final section is considering motivation as the drive that allows for learning with simulation games. In that section the importance of motivation for both games and learning is discussed, and specific engaging techniques are considered.

When discussing how *design-based research* can be conducted within a PhD thesis, Herrington and colleagues (2007) suggested that the literature review should result in preliminary principles to inform the design and development of the planned intervention. At the end of the iterative implementation of the intervention, these principles need to be refined and reinforced as a reusable, albeit contextualised product of the research. The principles derived from this literature review are presented in Chapter 4, and the subsequently revised principles as a result of this research are summarised in Chapter 9.

2.1 Learning

This section follows Davies and Mangan's (2006) idea that teaching should be built on theories about learning to derive pedagogical principles and following that, learning activities. Thus it considers first relevant theories and related principles, then it considers particular established and studied practices. The considered theories include research on cognitive load, variation theory and theories about cognitive learning and different assessment taxonomies as a model of learning.

2.1.1 Theories

Cognitive load theory deals with the efficient use of limited capacity of human short-term memory. Empirical research (Baddeley, 1992) has shown that the human brain's *working memory* is limited in the information it could operate with simultaneously. The process of learning allows for combination of concepts in *cognitive schemata*, which allows people to

process more complex constructs in their short-term memory (Derry, 1996). In this section, research on *worked examples*, learning material sequencing and imagination are considered from the perspective of *cognitive load theory*.

One very practical line of research stemming from the cognitive load approach comprises of experiments with *worked examples* as a learning technique (Sweller, 2006). This research has shown that learning from example problem solutions as a form of instruction is relatively easy for learners to grasp. Even before cognitive load theorists' research on worked examples, providing demonstrations and examples of different kinds have been a commonly used teaching technique. Although, I am not aware of any references to Sweller's research from game designers, in practice related approaches are heavily used in digital games, for example in game tutorials (Andersen et al., 2012; Shafer, 2012).

Cognitive load theory tries to draw conclusions on ways learning efficiency could be improved. In a series of experiments, Chandler and Sweller (1991) demonstrated that by integrating the learning materials in a coherent textual presentation is helpful, but only if the integration of information was crucial for its understanding. However, in that study, the authors did not make an attempt to distinguish between different types of learning (see below), and subsequently might have focused on *declarative knowledge* and without considering what effects the extraneous burden of integration had, and whether learners showed indications of developing procedural skills when engaged with text analysis and aggregation.

Coherence is closely connected to the issue of sequencing learning content. Naturally, learning of complex phenomena needs to happen gradually. Research shows that *whole-task sequencing* is more suitable than *part-task sequencing* (van Merriënboer and Sweller, 2005). In other words, this means that the gradual increase in complexity should be accomplished not through initially taking out dimensions and reintroducing them along the process, but rather through keeping these dimensions in the context and simplifying their dimensions. Possible ways to simplify is to reduce variability (or even fixate) certain aspects. Of course, when the specifics of the task do not allow whole-task sequencing, part-task sequencing is a second best choice which needs to be used to encourage learners to build cognitive schemata. Further research summarised by van Merriënboer & Sweller (2005) indicates that in order to improve contextual interference, and thus, schemata construction, the activities involving partial tasks should be mixed throughout the sequence, as opposed to blocks of task separated by topics. For the purposes of

memorising, Leahy and Sweller (2008) suggested that a guided effort to imagine the subject matter helps learners transfer information to long-term memory. Although they are also using self-reports to measure imagination, parallels between it on one side, and immersion and willingness to experience it on the other could be drawn.

On the other hand, in a meta analysis Hoffler and Leutner (2007) suggested that the animated cartoon graphics have a beneficial effect in learning of motor behaviour. Such findings could influence decisions about the way narrative content is delivered (like a particular *worked example*). A deeper analytic comparison between textual/vocal or animated content delivery (storytelling) could be considered for the purposes of the learning experience.

Marton, Tsui and colleagues (2004, p. 18) make reference to a study illustrating the variability of practice hypothesis, originally related to Schmidt's *schema theory* (1975), even though the original research by Schmidt addressed motor learning skills. It is worth mentioning that *phenomenographists* argued that they used the concept of schema, as it is being used by authors in the field of cognitive load theory and cognitive sciences in general. Taking such a schema-considering cognitive perspective, research in cognitive load theory also has indicated that adopting variation is a useful learning technique (Paas and van Merriënboer, 1994; van Merriënboer and Sweller, 2005). From a computer scientist's perspective, *phenomenography* could be considered to be an approach similar to usability studies for learning. Yet, such a comparison overlooks the influence learners (the users in the domain of learning) have over the objects that they learn, be it through the gap of understanding between teachers and learners, or the range of different understanding that develops as a result of learning (Åkerlind, 2005; Marton, 1992)

In the phenomenographic tradition the use of variation has been developed into an extensive pedagogical framework by Marton, and continued by Pang (2003; 2006). At the core of their pedagogical framework are the two notions of variation and invariance; the former essentially bringing the concepts into focal awareness, while the latter draws away the learner's attention from concepts that are less relevant to that particular learning experience (Marton et al., 2004). Marton explains that retraction of the invariant elements with transcendence and a situation of being taken for granted. He argues that this helps in focusing on what has been considered relevant. Similar attention-directing approaches have been widely used in the development of computer games. The original value that phenomenographic research has added, is the connection between focal awareness and

learning.

Marton and Pang (2006) outlined four conditions of learning that they claimed are necessary: contrast, separation, generalisation and fusion. *Contrast* stipulates that in order for a quality to be discerned, a mutually exclusive quality has to be experienced in parallel. *Separation* emphasises that certain dimension of variation can be discerned, only if other dimensions remain invariant or vary independently. *Generalisation* complements separation by focusing on the fact that discerning of a certain value in a dimension is easier when this value is kept constant and when other dimensions are changed. Finally, *fusion* stipulates that the interplay of two dimensions can only be appreciated when the two dimensions vary simultaneously. Practical implications of variation theory are discussed further in section 4.1.1 with focus on different perspectives on the main object of learning and the phenomenographic learning study as an exploratory research method.

In one of several experiments on worked examples, Quilici and Mayer (1996) examine the effects of variation of the examples used on learning. The benefits that they identify are not statistically significant. As an explanation, they theorise that in their case, the group that did not employ variation was able to identify the common (invariant) elements across the examples.

Based on an experiment, Ranzijn (1991) concluded that wider (dispersed) variation supports declarative knowledge, and less dispersed variation supports procedural knowledge. In his study, he also observed the retention effects in different forms of knowledge over time: declarative knowledge decreased, whereas procedural knowledge increased. Ranzijn did not explore the effects on conditional knowledge, which is critical for future expert practitioners – what graduate students are. Schilling and colleagues (2003) investigated the application of variation in the context of organisational learning. They argued that the introduction of certain variation needed to complement narrow specialisation had positive effects on organisational learning. As means of assessment, they use playing games as a practical activity which required the creative application of knowledge, and not just following instructions.

A recent development within the constructivist tradition has been around *threshold concepts*. It has resulted from an ongoing effort to simplify curricula across the UK educational system (Meyer & Land 2006). Research involving students, teachers and practitioners concluded that certain concepts (named *threshold concepts*) are both difficult and important to grasp, but once mastered, they make it easier for the learners to work out

related knowledge and its application. Thus, they are considered to be fundamental to one's understanding of a particular subject – embodying the metaphor of acting as a threshold within their corresponding discipline.

Meyer and Land (2006a) outlined five features that are typical for threshold concepts. Considered from the perspective of a person's understanding, threshold concepts are (i) transformative to one's understanding, in the sense that it involves both ontological and conceptual shift. Threshold concepts are also considered to be (ii) probably irreversible, once they are thoroughly understood. Other characteristics are that threshold concepts are (iii) integrative in that they help learners understand different concepts together; and (iv) bounded in their scope, in the sense that their comprehension affects the understanding of only particular knowledge domain. These characteristics are the reason they are often likened to conceptual gateways or portals. There is one last typical feature of threshold concepts, and it is that they are usually thought of as (v) *troublesome knowledge*, which is knowledge that is difficult to both understand and apply. Although not all troublesome knowledge is a threshold concept. If certain knowledge is troublesome, generally this is a strong indication that there might be a threshold concept related to it.

Despite the focus on threshold concepts, researchers on this topic have developed claims, broader than threshold concepts themselves. In a work that has become fundamental to the first of a series of books on threshold concepts (Meyer and Land, 2006b), David Perkins discusses the wider implications of constructivism on troublesome knowledge (Perkins, 2006). Without making claims for exclusivity Perkins, considers 4 types of troublesome knowledge and the possible constructivist approaches to address difficulties in acquiring it. The four types that Perkins considered are inert knowledge, ritual knowledge, conceptually difficult knowledge, and foreign knowledge. To address inert and ritual knowledge, the kind that is known, but difficult to apply, Perkins suggested problem-solving and problem-based learning. As a response to conceptually difficult knowledge, Perkins proposes inquiries from different perspectives that would challenge possible discrepancies between theory and observations. For foreign knowledge, the author puts forward perspective-taking and role-play. The suggested responses for the latter two types of knowledge arguably points towards the adoption of variation in perspectives and roles. In his first chapter (Perkins 2008) to the second book on threshold concepts (Land et al., 2008), Perkins takes a more generalistic approach to knowledge. He identifies three groups of knowledge – *possessive*, *performative* and *proactive*. The first one being the

knowledge of information; the second – knowledge of how to apply that information, the third type of knowledge – when to apply it. Perkins further writes "well-cultivated threshold concepts seem likely to foster proactive knowledge".

Considering the lengthy process to verify threshold concepts and the challenges of teaching them, once they are identified, teaching (in the sense of facilitating learners to develop an understanding) the threshold concept is not straightforward. Davies and Mangan (2006) outlined a path from theory to pedagogical principles, to learning activities that needs to be undertaken in order to enable efficient learning. Even though threshold concepts are not addressed in particular, Marton and Pang in their necessary conditions of learning (2006) illustrated the difficulty of teaching complex topics, that even if broken down into more manageable pieces, would still require dedicated time and attention to the fusion of those pieces. This is what makes representing threshold concepts for learning difficult. Yet, Davies and Mangan (2006) among other authors, suggested variation as a way to facilitate learning of threshold concepts. The approach has been further discussed, both by researchers studying threshold concepts and variation theory (Meyer et al., 2008; Pang and Meyer, 2010).

2.1.2 Assessment

Due to their interactivity, when compared to traditional instruction, games are experiential in nature. This leads to a consideration on whether they are a useful tool for a range of learning content and learning purposes. In order to examine this, I consider different theories about knowledge and learning. I review current research to examine advantages of games-based learning over other learning methods that have been identified.

One taxonomy, that is commonly referred to, is the one used by Sugrue (1995). It has been used when evaluating not just games, but also other forms of *experiential learning*, like *problem-based learning* (PBL). It considers three types of knowledge: *declarative* (know-what, or what Perkins calls possessive knowledge), *procedural* (know-how, corresponding to the type that Perkins describes performative knowledge) and *conditional* (Perkins' proactive knowledge, answering the question know-with, but also arguably know-when, and possibly others). Walker (2013) mapped the possessive-performative-proactive knowledge taxonomy to Rasmussen and colleagues's (1994) skill-rule-knowledge taxonomy. Currently, there is a shortage of strong evidence, but there are empirical indications that experiential approaches are more efficient when dealing with procedural and conditional knowledge, but not necessarily declarative. A meta-analysis of such

outcomes in PBL was conducted by (Gijbels et al., 2005). This fits in very well with Perkins' (2008) theoretic framework. Whereas, arguably there might be better ways to gain new possessive knowledge than experiential learning, applying this knowledge to a problem challenges the performative knowledge and informs proactive knowledge. Since Sugrue's (1995; Gijbels et al., 2005) taxonomy is older, and was used in the research by Gijbels and colleagues, it is the terminology that is used throughout this thesis. Yet, Perkins' work comes to indicate how threshold concepts address exactly conditional knowledge.

This perspective, based on Sugrue's work, can also be seen as a simplified representation of other more complex taxonomies of knowledge. One more faceted taxonomy that was developed by Biggs (1994). It considers seven types of knowledge, acknowledging that cognitive learning does not fall into a single scale. This classification considers the less tangible tacit and intuitive types of knowledge. According to Biggs tacit knowledge "is manifested by doing, and is not verbally accessible". Intuitive knowledge is felt and might develop before being expressed symbolically. Biggs suggested that these two types of knowledge fall into a hierarchy of abstraction towards declarative, theoretical and meta-theoretical. Here declarative is the widely understood formulation of facts; theoretical represents an abstraction from declarative; whereas metatheoretical is the level when scientific work around abstractions may lead to paradigm shifts, meaning that it could possibly introduce some sort of revision of previous knowledge.

Finally, Biggs considers the procedural and conditional types. Procedural is the knowledge of how things need to be done, formulating necessary event sequences or order of actions. Conditional knowledge involves making decisions based on the circumstances. In the author's words "conditional knowledge provides the metacognitive support to procedural knowledge". Whereas, declarative knowledge can be represented by *know-what* and procedural with *know-how*, it is less obvious what question is representitve for conditional knowledge. Barnsford & Schwartz (1999) refined an idea by (Broudy 1977) of the *know-with* type of knowledge.

In the context of crime prevention, Ekblom (2010, p. 29) proposed a knowledge categorisation grounded into the particular domain. When abstracted away from the specifics of crime prevention situations, and when looking beyond the types related to *factual* or *procedural knowledge*, the questions *know-what-works*, *know-who-to-involve*, *know-when-to-act*, *know-where* and *know-why* stand out. Yet, the corresponding knowledge is interrelated and I consider the types of knowledge represented by these

questions to be facets of *conditional knowledge*. Being able to respond to these questions in situations involving other people, adds a crucial layer to the challenge. This additional complexity in conditional knowledge, introduced by social interaction, is a key challenge in domains like project management and information security.

Bloom and colleagues' (1956) taxonomy of learning objectives is a well-established hierarchical framework for comparing outcomes of learning. In particular, my interest focuses on their taxonomy of the cognitive domain, where they identified different levels of learning objectives. An updated revision of the cognitive domain taxonomy has been suggested by Anderson, Krathwohl and colleagues (2000). The categories in the updated taxonomy incrementally range from remembering through comprehension, application, analysis and evaluation to creation. In this sequentiality, Bloom's framework is similar to the declarative part in Biggs's forms of knowing. This taxonomy has been used, for example, in a survey by Anderson and Lawton (1992), questioning simulation instructors about the particular tools that they use. It appears that a portfolio of assessment techniques needs to be used, in order to capture different forms of learning. Anderson and Shattuck (2012) point out that using a variety of measures is also characteristic to *design-based research*.

Assessing knowledge presents an inherent trade-off in the fact that more complex knowledge is also more difficult to assess (Price et al., 2010). One consequence of this is that *Multiple-Choice Question* (MCQ) tests can be marked automatically, but are able to capture learning only to a limited extent, considering that they allow only choice between a predefined set of answers. Commonly, these tests are used to assess declarative knowledge. There is a variety of other assessment techniques, but typically testing comprehension goes through exploring learner's thinking process, either by written (like *essays*) or oral (for example *viva*) reasoning. On the other hand, testing the ability to apply what has been learned goes through learners actually demonstrating application, in either a real or simulated context. As successful forms of assessment are being widely used, essay writing and task simulations are considered here in more detail.

Essay or free-text writing is deeply rooted in culture. As such there have been many efforts made to automate the comprehension and assessment of essays. Such automation work (Butcher and Jordan, 2010; Shermis and Burstein, 2002) has had a huge impact with the adoption of e-Rater for GRE tests in USA. However, Powers (2002) demonstrates that the e-Rater assessment systems is not too difficult to exploit and mislead. Examining Biggs's

Figure 3: A schematic visualisation of the SOLO taxonomy, as illustrated by (Biggs and Tang, 2007)

(1982) Structure of Observed Learning Outcome (SOLO), a recent taxonomy intended for the evaluation of written text could provide greater insights as to why automated text assessment is so challenging. SOLO (illustrated in Figure 3) examines the structure of written answers, defining levels of understanding based on the structures that have been used in the content. These range from pre-structural, through uni-structural, multi-structural and relational to extended abstract. This discourse structure analysis is challenging even for experienced practitioners. As discussed in the next section, similar limitations of machine interpretation of free-form written language are also being encountered in automated conversational agents.

Although, widely adopted in a number of fields (like pilot training), task simulations also have their limitations. The human interface of a complex piece of machinery is deterministic, and thus, could be simulated with accuracy that would guarantee transferability to its real-world use. However, simulations of interpersonal relationships are less similar to original practice, and more challenging to implement. The work of Reeves and Malone (2007) shows that there is evidence of successful leadership learning in virtual worlds, but Anderson and Lawton (2009, p. 205) noted that *"simulations' time utilisation relative to other pedagogies often raises questions regarding its efficiency for imparting learning"* along with the observation that research in business simulations often

lacked the necessary scientific rigour.

All this comes to expose some of the reason for the limited success of serious games and related experiential learning methods. The complexity of the problem comes from the combination of subjectivity of assessment for more complex forms of learning and the limitations of efficiency of games for learning of declarative knowledge when compared to traditional learning methods like reading a textbook.

Due to its importance for learning, feedback is typically part of an assessment, be it formative or summative. Several different authors, like Walker (2009) and Moreno (2004), make strong cases of the benefit of explanatory vs corrective feedback. In other words, it is important not only to indicate to learners when they are right or wrong, but also what was the reasoning in the assessment, so they can reconsider their answers (and hopefully their way of thinking) in the future.

2.1.3 Practices

Given the limitations of automated and objective assessment, there is strong support for situated learning at the workplace. A selection of this tradition will be considered here, namely *learning by doing*, *communities of practice* and *knowledge-sharing*.

A common contemporary understanding of expertise is the one of situated cognition (Brown et al., 1989). The main idea is that learning happens best in a practical context and through *cognitive apprenticeship*, in other words through learners trying to mimic the way professionals think. A key principle in situated cognition is *learning by doing*, closing the gap necessary for knowledge transfer from learning to applying in practice. Learning by doing is considered especially useful when certain degree of tacit knowledge and the need to externalise it are involved.

Several research communities have studied different perspectives on what is being learned. Here, three overlapping models of learning are considered: (i) Marton, Tsui and colleagues (2004) developed their perspectives in the context of their *object of learning*, (ii) Laurillard (2001) puts forward her *conversational framework*, and (iii) Winn (2008) proposed his *Design-Play-Experience* framework for serious games. All of these (see Figure 4) provide very elaborate discussion of different perspectives towards the object of learning and the transformation it takes through the stages of teaching by teachers, respectively learning it by learners.

Marton's perspective focuses on the evolution of the object itself: how is it *intended* by the

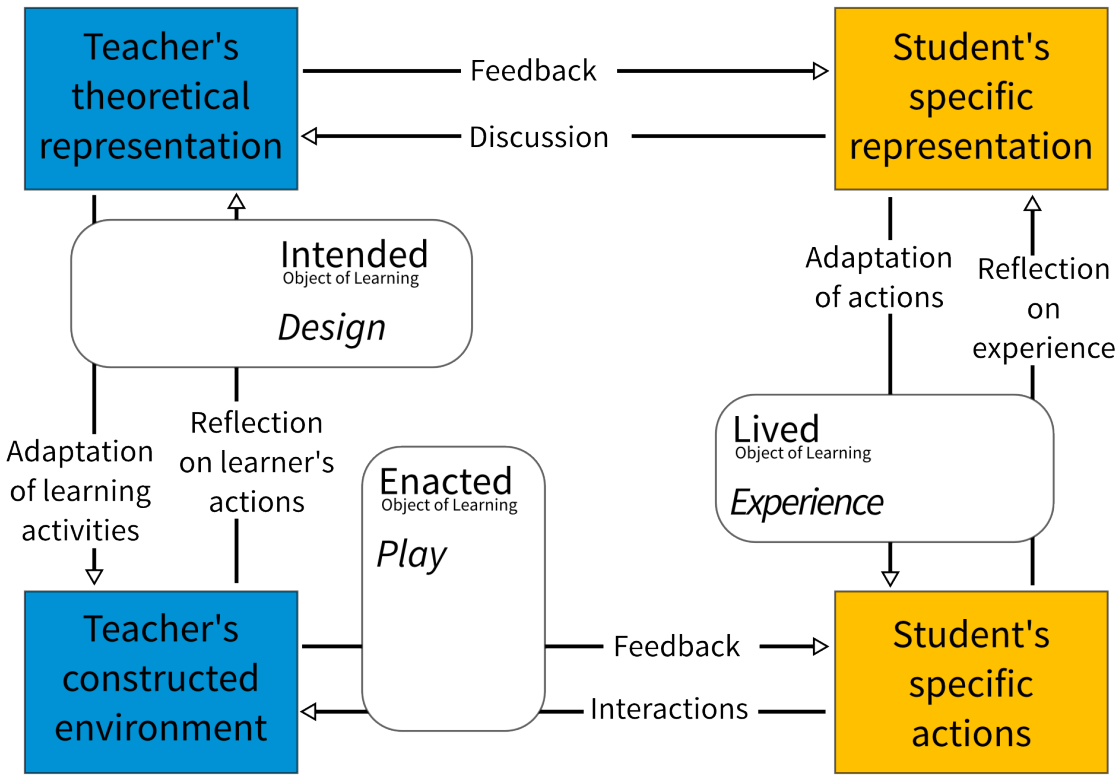


Figure 4: The representation of the manifestations of the object of learning in the three frameworks, based on Laurillard's simple (single-learner) model, as described in (Laurillard, 2001). The four boxes in the corners represent the teacher (left) and learner (right). The manifested forms are overlayed in white boxes, the DPE concepts are in italic.

teacher, how is it *enacted* (meaning what happens) during the learning activity and how is it *lived* (perceived) by the learner. Thus, these three manifestations of the object of leaning represent the transformation of what is intended to be learned towards what the learner perceives of the actual learning experience. These three manifestations exactly correspond to Winn's take on serious games. Marton's intended object of learning is the design of the learning activity in Winn's terms, or serious game if that is the representation of choice. Winn's play is how the object of learning is enacted in serious games. And finally, the lived object of learning in variation theory which corresponds to the way the player perceives the experience. However, Winn expands his framework further by breaking down each of the three manifestations into four layers, that he considers relevant to serious games. These layers are learning, storytelling, gameplay, and user experience. Naturally, the layer that most closely relates to Marton's perspective should be learning. Winn's matches content and pedagogy to design, teaching to play, and learning to experience. However, this formulation does not capture the complexity of the interaction, as both the object of learning and single-layered DPE framework does. Similarly, the storytelling layer leaves some ambiguity in separating narrative on the design side, storytelling on the play side, and story on the experience side. Yet, the gameplay and user experience layers more clearly

reflect the relevance when compared to the object of learning. The designer's intent (relate to intended object of learning) is captured in mechanics and user interface for each of the layers correspondingly. Similarly, dynamics and interactivity reflect the process of the learner engaging with the learning process of the serious game, as well as learner affect and engagement reflect the subjective experience they take away from this activity.

Laurillard's approach also considers the exchange between the teacher and the learner, but it more clearly illustrates the gap between the teacher's intentions and the outcome with the learner. Each of the two perspectives is represented at two levels – cognitive representation in both teacher and student. It is observed – how does the teacher construct the learning environment, and how does the learner act within it. Laurillard also represents the information exchange between each of these different forms of the learned object, as illustrated in Figure 4.

These three different perspectives of generally overlapping concepts provide ground for a very elaborate discussion. Due to the importance of this discussion and research for this thesis, it is elaborated in further detail as part of the overview of methodologies that were adopted in Chapter 4.

Professional practice comes with a recognition of the fact that part of our knowledge is tacit. Collins (2001) identified several types of tacit knowledge, ranging from ones that are intentionally not made explicit, to ones that people in their possession are not aware of. There are a number of techniques that have been suggested to address the transfer of tacit knowledge, notably ones that suggest collocation and mutual observations between expert and apprentice (Nonaka & Takeuchi 1995). The current practice of tackling *tacit knowledge* is to support collaboration and communication as means of knowledge externalisation (the process of turning tacit into *explicit knowledge*) and transfer.

Communities of practice (CoP) are professional groupings where practitioners exchange experiences and advice. These communities are where professional practices and professional language are being developed. Lave and Wenger (1991), the original researchers on communities of practice, suggest that the process of developing an identity as part of the community is essential for the transition from a novice to an expert. With the advent of networking technology, communities of practice have partially shifted to digital media. This has led to the expansion and globalisation of many communities of practice. Due to the changes that are inherent in such a shift, virtual communities of practice have two major benefits. On one hand, they provide an opportunity to gain better understanding

of communities of practice on one hand. On the other, they allow for the introduction of a number of new technologies that ultimately change the way community members interact. Trust, peer learning and role-models are discussed in the context of communities of practice.

Kanawattanachai and Yoo (2002) conducted a study showing that given a comparable start, virtual teams that manage to build trust perform better than ones with lower levels of trust. They also developed an argument that virtual teams rely more on the cognitive than on the affective element of trust. They call cognitive the rationally motivated trust, whereas the affective is the one that has been emotionally-driven. A study by Fang and Chiu (2009) confirmed that trust in fairness of the socio-economic system is an important factor in determining knowledge-sharing (in other words contributing) behaviour of community members. Fang and Chiu also elaborated on the existence of two notably different forms of trust in teams: i) trust in the group as a whole, but also ii) trust in a particular member. These forms of trust can be seen as built in a series of transactions. The former has been discussed above, whereas the latter is addressed further below. Riegelsberger and colleagues (2005) proposed a transactional model of trust, which could possibly be adapted to be used in a virtual community.

In our work with Malheiros we (2011) considered trust in the context of serious games. We identified a factor that plays a key role in building trust between users is transparency about how data collected in the game is being used. Users are very sensitive when the system collected data for their behaviour and who has access to what is collected. One of their recommendations is that in order to encourage trust, a serious game needs to make a clear distinction between learning and assessment modes, allowing for unobserved practice in learning mode.

Peer-learning has been traditionally used in various educational systems across the world. A number of studies confirmed the benefits of peer-learning, like King and colleagues (1998) reported improved efficiency with the adoption of peer-tutoring. Peer-learning can be valuable when situation-specific knowledge is involved (which is conditional knowledge in Sugrue's classification), when it is difficult to prescribe general rules.

In the context of Gijbels's findings, peer-learning is not only a good match for communities of practice, but also suggests the potential for peer assessment, which could give the additional benefit of written assessment that is validated by other people. Of course, peer-assessment raises the issues of validity that would otherwise be guaranteed by the

expertise of the teacher. Ploegh, Tillema and Segers (2009) conducted a survey with teachers that use peer assessment and found that the quality of peer assessment is still in the hands of the teacher, and hugely depends on the teacher's ability to organise a quality appraisal process.

The personal example of role-models has been recognised around the world as successful learning practice. It is popular as an influencing factor in child education, but research shows that the affiliated notion of self-concept is a factor throughout one's entire life, as examined for example by Gibson (2003). In his study in particular, Gibson showed that with experience people start to identify not only positive, but also negative role models and learn to pick the influences they adapt from their role models. Shuval and Adler (1980) also emphasised the processes of *active identification* and *rejection* behaviours learners adopt. Yet these authors make a step further and note the existence of a parallel process of *inactive observation*, which serves as a continuous validation of the self-concept through ununtentionally absorbing behaviours from others. Arguably, there is some resemblance between *active identification* and *worked examples* (see previous section). In both cases, actual performance is used as a form of self-reassessment. Whereas in the latter a particular problem-solving exercise is being considered, the former focuses on the perceived identity of the person performing it.

Social platforms like Wikipedia, Answers.com and Delicious are known for successfully crowd-sourcing knowledge. Similar effects can be observed in community-driven platforms that are more dedicated to particular domains or topics. Sermo, a professional network for US-based physicians (Bray et al., 2007), is an example of living knowledge ecosystems utilising Web 2.0 tools. Sermo features surveying, tagging, discussions and rating among other techniques. The collectively constructed practical knowledge, developed within that community is a product that has been considered to be of commercial value, which itself is an indirect evidence of data quality. Such a perspective considers knowledge as being dynamic and therefore considers people as part of the knowledge ecosystem. There is only limited research on specific attempts to integrate knowledge sharing in games, like my work reported in a separate publication (Ruskov et al., 2010), despite the potential added value, similar to the one mentioned for peer assessment. However, in more general online learning, knowledge sharing is what typically happens in *inquiry learning* (IL) and *personal learning environments* (PLE), as explained below.

A key advantage of communities of practice and knowledge ecosystems is their flexibility and their suitability to provide answers to specific questions and suggest adaptations of these answers to the specific context of the person enquiring. This flexibility makes IL more accessible for community members that are making the inquiries. Thus, knowledge sharing can be used both for assessment and externalisation in a peer-learning process.

Another way to look at knowledge sharing – as a form of written reflections – has been recognised as a valuable learning tool (Moon, 2001). It has been claimed that it benefits not only the viewers, but also the contributors, who externalise and thus reinforce their knowledge. Moon, among others, emphasised the importance of human guiding in the reflection process. This falls in line with the evidence that in educational games, the process of reflection after the game (Hays, 2005) and the breakdown of tasks (Reeves & Malone 2007) are essential and an argument supporting the advantages of shorter written lessons can be developed. Short texts have two advantages over long ones. The first, is that shorter contributions effectively reduce the threshold for participation, and the second – is that short texts may be easier to transfer between people, as their shortness may be related to less contextualisation.

When considering life-long and personalised learning, a related and more neutral perspective is being taken by researchers (Attwell, 2007), advocating a *personal learning environment* (PLE) approach. According to Attwell, one key characteristic of PLEs is the fact that they typically integrate different online services. Some researchers tend to view PLEs more broadly, including also classical learning materials and tools. The idea behind a PLE also acknowledges the role of individuals in organising their own learning in different contexts and situations (Wilson et al., 2006). Moreover, it takes into consideration the fact that learning support and new knowledge are generally not delivered to the individual by a single source. Arguably, this is a form of acknowledgement for the importance of role models in the learning environments. In line with the aforementioned reference to social platforms, such platforms and social awareness streams (most notably *Twitter*) in particular have been recognised as one of the key elements of personal learning environments (Kompen et al., 2009).

All in all, the ideas of learning by doing represent a criticism of the way formal education is traditionally delivered. An argument was made (Wenger et al., 2002) that in the academic environment, a superficial community of practice emerges implicitly and displaces the practitioner community that novices need to join in order to develop their own expertise.

Experiential learning helps academic communities of practice refocus and closer match the original professional ones.

2.2 Games

As with all games, entertainment also needs to be an intrinsic property of any simulation game. Here a widespread model of game design is considered. It has been commonly accepted among game designers, but developed in depth in writings from practitioners in *table-top role-playing games* (Edwards, 2001) and digital games (Lindley, 2003). The representation of the model features the three core components of simulation, narrative and game. In this text, the model will be called the *game design triangle*. Each of these components could hugely vary, depending on the purposes of the current product, but here their application in the context of educational simulation games is discussed.

2.2.1 Definitions

Research in domains relevant to games for learning is very diverse, and correspondingly there is an extremely broad range of related terms. *Serious games* have emerged as the dominant term (de Freitas 2006; Susi et al 2007; Ulicsak & Wright 2010). These are games that in contrast to the established idea of a game, have not been designed for the primary purpose of entertainment. This is a broad definition, and although it includes for example *advert-games*, the term is mostly used for serious games for learning and this is the way it will be used throughout this thesis. Researchers from various communities and backgrounds use the terms *educational* (Kebritchi & Hirumi 2008; Kiili & Lainema 2008) or *instructional* (Hays 2005; Malouf 1988) *games* which distinguishes the different approaches used. This variance of naming stems from different communities and backgrounds, and correspondingly to different approaches. Throughout this thesis I consider any claims or results in one research area, using specific terminology to be transferable to the others. When discussing relevant literature, effort has been put to preserve the original terminology.

Research into business simulations and games has a long history. For reviews, see the work of Sauve (2007) and Maier and Größler (2000). A number of authors have put forward definitions of simulations and games, and have attempted to distinguish between these transcending terms. One common observation made by various authors (Aldrich, 2009a; Sauvé et al., 2007), is that each of the aforementioned categories focuses on a particular set of features that are not essential, albeit often present in the other categories.

Sauve and colleagues (2007), among others, conducted a review of the essential elements of games. In that review they reduced the definition to five distinct features. These are (i) the existence of player or players, (ii) conflict, (iii) rules, (iv) predetermined goal and (v) artificial nature. In the context of educational games, the authors introduced (vi) the pedagogical nature of the game as a sixth feature.

In their review, Sauve and colleagues (2007) also summarised some essential attributes of simulations used for learning. They identifed four key attributes: (i) a model of reality defined as a system; (ii) a dynamic model; (iii) a simplified model; and (iv) a model that has fidelity, accuracy and validity. Here too, the authors introduced a pedagogical attribute: (v) the simulation should directly address the learning objectives. In a somewhat more unifying perspective, Aldrich (2009) argued that "simulations use rigorously structured scenarios carefully designed to develop specific competencies that can be directly transferred into the real world."

While games and simulations are generally viewed as distinct technologies as indicated earlier, some systems combine elements of both – typical features of simulations (for example, a predominantly realistic game environment) with some features of games (like competition, user-friendliness, and others). These technologies are commonly called *simulation games*.

Further two related categories have been discussed in the literature: interactive storytelling and virtual worlds. *Interactive storytelling* (sometimes called digital storytelling) is widely claimed to have emerged from games (Mateas 2001; Crawford, 2004; Mateas, 2001). On the other hand, Riedl, Stern and Dini (2006) elaborated on the distinction between simulations and interactive narratives. Interactive storytelling has been contrasted with both games (Crawford, 2004) and simulations (Riedl et al., 2006) in its emphasis on a pre-designed story. In the core of a definition of interactive storytelling is the conflict between interactivity and a predefined plot. Authors in the field (see the previously quoted as examples) developed arguments of possible controlled interactivity that would still maintain the initially intended narrative. When talking about educational technologies, focusing this narrative on the learning material can become a strong learning support technique, as has been demonstrated by Marsella and colleagues (2000).

In his comparison between games and simulations, Aldrich (2009) also included *virtual worlds* as a distinct third category beyond games and simulations. He underlined the absence of a specific goal as a distinctive characteristic of virtual worlds. He also indicated

that communication is a much more central feature for virtual worlds, as opposed to the other environments discussed.

2.2.2 Simulation

In order to provide coherence to their story world, games commonly employ some form of simulation. As a form of experiential learning, the simulation in a game is where learners get the opportunity to go through a realistic experience that would benefit their practice. In order to achieve this, the simulation component needs to allow for realistic problem-solving. Two widely used models of experiential learning are *problem-based learning* (PBL) and *inquiry learning* (IL). Each of them have emerged as a way to learn in professional practice.

PBL is a task-driven approach to learning, which puts learners in the realistic situation of having to solve a particular business problem. When the simplified representation of this problem is realistic enough, it allows learners to transfer and apply to practice what they have learned.

From a slightly different perspective IL also aims at solving problems, but is focused on the process of filling in gaps of the student's understanding by means of questions and answers. In a typical IL process, learners identify open questions while working on their tasks and initiate discussions and inquiries to answer them.

The survey-based study of *massive multiplayer online role-playing games* (MMORPG) that Reeves and Malone (2007) conducted can be considered to be an illustration of both PBL and IL. Players in these MMORPGs have to organise themselves for their collective gaming activities, and they have to discuss and negotiate their tasks. The study provides evidence that these players can acquire leadership skills that are transferable to real-life situations. The authors also built an argument for the benefits of breaking down the tasks into manageable pieces. They claimed that this allows players to develop higher flexibility in assignment of leaders, and thus, according to the authors, to the development of better leadership skills among those employees that play such games. This approach of breaking down information into smaller chunks is discussed again in Section 2.3, in the context of research on motivation.

Experiential learning techniques, including PBL and IL have been criticised with a recurring argument from several studies that have discussed the shortcomings of unguided learning (Kirschner et al., 2006), (Mayer, 2004). However, such interpretations are

oversimplified, because the studies demonstrated that minimal guidance or the lack of it, during learning appears to be less efficient when compared with methods with guidance. Such conclusions can be seen in the number of responses to Kirschner's article (Hmelo-Silver et al., 2007; Schmidt et al., 2007). Therefore, the collected evidence could be interpreted as arguments in favour of scaffolding, but not necessarily opposing PBL and IL. Such interpretation is reinforced by the meta-analysis of games by (Hays, 2005). He concluded in favour of the importance of debriefing and feedback when using instructional games.

A dedicated body of ITS research focused on the relationships between feedback and students' affective state and outcome (Boyer et al 2008; Del Soldato & Du Boulay 1995). Some of it emphasised the value of positive feedback (Barrow et al., 2008; Di Eugenio et al., 2005) in line with established practices, as this will be explained in more detail in Section 2.3.

2.2.3 Narrative

The second key component of the game design model is narrative. As already discussed in the context of cognitive load in the previous section, when storytelling supports people's imagination, it also supports memorisation. Storytelling can also be used to relate the game and the targeted learning context. There is now a well-established research tradition in a field that encompasses the narrative side of computer games, and is broadly called *interactive and digital storytelling* (IDS). Several developments from this field are considered here.

A branch of IDS focused on attempts to develop dialogue-based *intelligent tutoring systems* (ITS). These attempts have attracted significant attention in the ITS community (Kerly & Bull 2006; Litman & Forbes-Riley 2006). Yet, applicability of such attempts in complex learning have a number of technical issues to resolve and only partial examinations have been conducted.

A number of scientists advocated for the literary approach to interactive storytelling that was elaborated by Murray (1998). This led to an attempt to design a balance between plot and interactivity. Such developments assume dependence on a predefined plot, and put a constraint on players not to deviate from the author's intended design for a richer experience.

A different – although related – approach has been taken by other researchers. Aylett and

colleagues (2007), for example, advocated for an approach they called *emergent narrative*, focusing on the characters as the leading generators of interactive narrative. This approach had been influenced by RPGs that have been widely recognised as delivering successful storytelling experiences that arise as part of the player's interaction. Various aspects from table-top and live action RPGs have been adopted for *massive multiplayer online games* (Tychsen et al., 2006) and *alternate reality games* (McGonigal et al., 2006).

A key challenge in digital game design is the creation of believable non-player characters (NPC). Early attempts in interactive storytelling, like (Iurgel, 2006; Mateas and Stern, 2004) attempted to derive character behaviour from the intended dramatic plot. A different approach, considered by a number of authors is character-driven storytelling. This approach is often referred to as Egrian, after its author Lajos Egri (2007), as opposed to the Aristotelian idea of a plot that dominates the characters. It is also closely related to the concept of emergent narrative. In essence, these approaches showed how role-play can be expanded into storytelling, so that multiplayer games could benefit from a more engaging interaction. The work of Marsella and colleagues (Si et al., 2009), as well as Aldrich's *vLeader* (2009b), are successful examples for a practical implementation.

It could be argued that with games in general, the inherent complexities of dialogue-based interaction require complex modelling of the specific context of the domain. Such an approach is taken in task-based dialogue management systems, for example the work of Xu and Rudnickly (2000). In an earlier paper (Ruskov and Ruskov, 2006) I proposed a design for a dialogue-based simulated negotiation game that could potentially allow a deeper level of automated discourse processing and interaction in the restricted domain of multi-dimensional negotiations. The negotiation pattern is centred around a controlled language, which is mapped onto a game-theoretic space as a part of a wider human-scaffolded exercise (see Section 2.1 for a discussion of *cognitive scaffolding*).

2.2.4 Ludology

The third and final component of the game design model is the ludic (or game) element. For the purposes of the current research, two particular aspects of this component are considered – personalisation and variation in games.

Games, as commercially distributed products have been developed with heavy consideration of the player's engagement. Game designers have used many of the widely accepted motivational techniques (see discussion in Section 2.3), but their pragmatic

approach towards personalisation arguably appeared more successful than mainstream attempts in the scientific community. As a motivational factor, personalisation will be brought up again in the next section, but its realisation in games as a design feature will be discussed here. Personalisation in games has found expression naturally through distinct market segments and genres, such as sports games, business simulations or casual games. In many games, various options for players to customise their experience have been provided: commonly used features include avatars, props and traits. These are features that players can choose, find or develop throughout their experience. Digital games so far have only made limited use of intelligent personalisation and recommender systems, one notable exception is the work of Tanenbaum and Bizzocchi (2009). As discussed further in the corresponding section on *personalisation*, this is a very different from the approach taken by the research community that has been developing *intelligent tutoring systems* (ITS). It could be argued that in contrast to mainstream ITS researchers, game designers have acknowledged the diversity of motivational factors for players and the fact that some players might not appreciate certain types of personalisation (Tychsen et al., 2006). Furthermore, it could be argued that the way game designers have adopted personalisation has similarities with the notion of an open learner model (OLM), which has been adopted in some ITS. This is also discussed further in see Section 4.1. OLM explicitly involves users in personalisation-related decisions, as entertainment games usually do.

Jesper Juul drew a connection between variation over time in games and continuous game attractiveness (Juul, 2007). Essentially, he elaborates on the attractiveness of gradually increasing the scope of variation with the progression of a game. Despite the fact that he intuitively used the notion of variation, the meaning that he employed and his conclusions could be possibly connected to the more rigorous variation theory of learning (Marton et al., 2004). Whereas Juul talks about attractiveness, learning theoreticians discuss focal awareness. If Juul's claims prove to be more universally valid, this might serve as another strong argument in favour of the use of variation in learning.

2.3 Engagement

Engagement and motivation are well known drivers of behaviour. The fact that they are engaging (Wright 2009) is one of the core reasons for the interest in them. However, as will be discussed in this section, human motivation is a complex topic. Certain engagement techniques may work positively in some situations, and be demotivating in others. Different people could even react to the same feature in different ways when focusing them

in different contexts. In this section, approaches towards considering motivation, and subsequently specific techniques that aim to capture players' interest are discussed.

2.3.1 Motivation

For the purpose of establishing a theoretical framework when discussing motivation, several research efforts are considered here. The section starts with the popular *carrot-and-stick* approach and discusses theories that are more narrowly contextualised about the flow and motive that drive people to play online games.

In classical behavioural and social psychology, the combination of reward and punishment were studied extensively and found to be important ways to motivate people. This carrot-and-stick approach has been widely used in business, in the form of bonuses for example, career advancement opportunities and/or financial penalties. It is now widely accepted that positive incentives, and positive feedback in particular, are more efficient than punishment when it comes to motivating people.

However, providing external incentives is not a straightforward way to motivate people – these work in conjunction with the person's own motivational factors. In particular, a number of studies have converged to the conclusion that extrinsic rewards could possibly undermine intrinsic motivation (Deci et al., 1999). Such research suggested that people interested in motivating others in a business context should look for ways to scaffold intrinsic motivation – the person's own reasons to engage in a particular activity.

Cameron, Banko and Pierce (2001) questioned these findings and suggested that negative effects can be worked around. They conducted a meta-analysis that concluded that rewards do not generally undermine motivation to perform a specific task. They classified rewards according to the interest their participants had. On low-interest tasks they found that rewards reinforced autonomous intrinsic motivation. When interest in the task is high, verbal rewards had positive effects on motivation and self-reported interest. They found negative effects on high-interest tasks for tangible and expected rewards, but also on inadequate rewards – those that are loosely tied to the level of their performance. These findings actually reinforced the critique of numeric rewards on high-interest tasks and the way they are commonly introduced, but indicated that when applied carefully, extrinsic reinforcement could actually support and complement intrinsic motivation.

Generally, intrinsic motivation is connected to personal needs like growth, challenge, curiosity, autonomy, direction, mastery (Pink, 2011). A key theory dealing with intrinsic

and extrinsic motivation, *self-determination* theory, distinguished the particular need for competence, which had been suggested to be met through verbal reinforcement and positive feedback (Deci, 1971).

The theory of *flow* takes another perspective of the need for competence and mastery. It was introduced by Csikszentmihalyi (1991), and is based around the idea that optimal experience is a middle ground between boredom and frustration. Csikszentmihalyi argued – and there is supporting evidence – that a good balance of perceived challenges and perceived skills is a necessary condition for someone to experience flow. Goal-setting and immediate performance feedback have also been identified to support flow experiences (Csikszentmihalyi & LeFevre, 1989), which is congruent with self-determination (see above). Being in a state of flow often includes a sense of being in control and – unlike self-determination – experiencing time distortions.

Another of the motivational drives referenced by Pink is *autonomy*. Various authors, like (Greenberg, 1992; Zuckerman et al., 1978) found that giving people the opportunity to choose between several options (that is autonomy) served as a strong motivating factor. However, as a motivating factor, it should be used cautiously as too much autonomy can overwhelm people and inhibit performance – a phenomenon popularly known as paralysis-by-analysis (Schwartz and Ward, 2004). In the context of cognitive load theory it could be predicted that a reasonable guideline for a number of alternatives would be an upper limit of about seven.

While all this considers general motivation from a behavioural point of view, there have been several attempts to classify game players according to the motivational factors that

Achievement	Social	Immersion
Advancement Progress, Power, Accumulation, Status	**Socializing** Casual Chat, Helping Others, Making Friends	**Discovery** Exploration, Lore, Finding Hidden Things
Mechanics Numbers, Optimization, Templating, Analysis	**Relationship** Personal, Self-Disclosure, Find and Give Support	**Role-Playing** Story Line, Character History, Roles, Fantasy
Competition Challenging Others, Provocation, Domination	**Teamwork** Collaboration, Groups, Group Achievements	**Customization** Appearances, Accessories, Style, Color Schemes
		Escapism Relax, Escape from RL, Avoid RL Problems

Figure 5: Motivations for play as shown in Yee (2005)

make them play. A bespoke early taxonomy by Richard Bartle (1996) suggested four types of players, and a questionnaire was developed to classify them. Starting with this questionnaire, Yee (2005) conducted a more extensive investigation and based on factor analysis on participants' responses in his survey, concluded that in reality the different motivational factors are not exclusive to one another. His analysis also concluded that the motivational factors could be grouped in three groups, each with its own category of sub-components (see Figure 5). He names the three top-level factors to be achievement, immersion and social motivation. Each of these factors are the aggregation of yet other, loosely related factors. The *social* motivation unites the needs for socialisation, relationship building and teamwork. The *achievement* motivation represents the needs to make progress, to compete and to understand the underlying rules. The *immersion* motivation combines discovery, role-play, customisation and escapism. Another interpretation of this factor could be *fantasy*, as all subcategories seem to involve some form of creative play, in the sense that something is created, be it in the game world, or as a mere imagination. For the reason that these labels were assigned to the results of factor analysis, the names should not be considered exact descriptions of the underlying phenomena. For example, while the sub-factors for achievement have an average factor loading of more than 0.75, for immersion the average is weaker and close to 0.65. More detailed insights are available in the original publication (Yee, 2005). Yee considered the labels of achiever and explorer to loosely correspond to whether the player is extrinsically motivated or intrinsically motivated, respectively.

Yee's research covered 3200 participants and drew conclusions about the demographics of play. His findings indicated that close to 40% of respondents were primarily motivated by immersion, but the other two major groups of motivating factors differ across age and gender. For example, according to this data typically men tended to be more engaged by achievement, whereas women in social motivations. This leads to the need to consider the demographics of the target audience when designing a serious game.

Finally, rather than generic reasons why people do things, Yee's research was grounded in personal experiences of players of online games. This closely connected it to the intended experience, which the current research effort aimed to deliver.

O'Brien and Toms (2008) developed a psychological model of engagement with different types of software and other media. The applications they considered are four types: video games and learning via webcasting alongside web shopping and searching the web. They

drew a clear distinction between factors that contributed at the point of engagement when users were drawn to use the tool; factors that kept them engaged and factors that contributed to disengagement. In the results to their exploratory interviews they report cases of participants engaged by social motivation, aesthetics of the interface, and personal interest. These can be mapped to Yee's factors which motivated players to play multiplayer games. Social motivation is a clear match of major factors in both models. Aesthetics can be mapped to various factors of immersion, and various forms of interest could be seen as a combination of achievement and discovery.

O'Brien and Toms also considered factors that were attributed to continued engagement on one hand, and disengagement on the other. These go beyond the motivational factors that are necessary to become engaged. Such factors are related to the perceived response a user gets from the system. O'Brien and Toms listed the following factors: appeal (both aesthetic and sensory), attention, awareness, control, interactivity, novelty, challenge, feedback, interest, positive affect for engagement and usability, challenge, positive and negative affect and perceived time for disengagement. Among those factors are the ones that are extrinsic to the player. The majority of these factors can be addressed with engagement techniques described in the next section.

Since motivation is so important for learning, it would be beneficial to reinforce it with further engagement techniques, as typical to serious games.

2.3.2 Engagement Techniques

Beyond the broader theories there have been empirical findings about particular engagement techniques. Here research about persuasion, personalisation and information chunking is summarised.

One factor that could be potentially considered to complement motivation is persuasion – the process of convincing someone towards a change of behaviour or attitude. Chaiken (1987) argued that people that are less intrinsically motivated are more susceptible to less cognitively demanding persuasive techniques. A number of technology-enabled opportunities to persuade, have been identified by Fogg (2009, 2002) in what he termed *captology*, or the study of *persuasive technologies*. Computer games traditionally feature a number of persuasive techniques identified by Fogg's framework. Moreover, Will Wright, one of the leading game design practitioners, also developed an argument for the motivational value of games (Wright, 2009). However, in the context of game design for

learning, it might be beneficial to approach achievement-driven learners with some form of simplified persuasion techniques, beyond the game scores that are commonly present. This needs to happen without engaging their full attention, so their awareness can be focused on learning and reflection.

As already discussed (see previous section), many computer games provide features to allow players to personalise their gaming experience. Personalisation has also been heavily used in technology-enhanced learning (TEL). Researchers in the field of intelligent tutoring systems (ITS) have employed various *recommender system* techniques to adapt the content they deliver. Partly because of the ease of access to interaction logs, data-driven approaches were widely adopted. Although some connections to motivational theories like self-determination could be identified, the research is mostly empirically-driven. For example, Cocea and Weibelzahl (2009) make assumptions about the motivation of learners, based on the time they spend on a particular section of the learning materials. Their intention is to adapt the learning process to the motivational symptoms they get from the learner. The fact that the approach is driven by system data, makes it challenging to validate their assumptions about the cognitive processes that underlie their observations. This becomes even more challenging if such an approach is to be attempted when more complex forms of interaction like ones of the type present in games. To accommodate the complexities related to personalisation, a number of ITS adopt probabilistic (Conati et al., 2002) or case-based modelling (Cocea and Magoulas, 2009; Hulpuş et al., 2010) of the user. However, the need to conduct validation studies not only of the implementation, as is already being done (Cocea, 2006), but also of the underlying psychological assumptions. With the continuous design of new systems, both these systems and the claims they make about transferability of the theories they adopt need to be empirically validated.

In online games, learner models blend with player characters. The basic idea of opening such a model to the learner is that in this way they get full control over their profile within the game, including the presumptions the system makes about them. This is used as a form of personalisation (also see 2.1.1). This model could be partially filled by various assessment methods, like Kolb's learning styles have been widely used (Stash et al., 2004), but in OLM, the learner has the freedom to edit this profile at anytime. In their work Bull and colleagues, on one side (2008) and Tanimoto on the other (2005), suggested overviews on potential dimensions of OLM.

Beyond persuasion and personalisation, a third technique that is known to work well on motivation to think is *information chunking*, meaning breakdown of information in manageable pieces (Petty et al., 2001). Wood (1982) observed that people get motivated about a specific goal when they are better informed. In her experiment, Wood also demonstrated that people who are less knowledgeable about the topic, are more susceptible to opinion influences, when compared to more knowledgeable people. An argument could be made that this has resemblance to the recommendation of the joint survey made by *IBM* and *Seriosity* on leadership skills in MMORPGs (Reeves and Malone, 2007).

2.4 Games for Learning

Despite the intensive work in the area of educational games there are still a number of fundamental open questions. There is already solid evidence of the benefits of playing computer games for certain learning domains at an early age (Prensky, 2003). However, possibly due to the complexities involved in real-life, few games have proved effective at the level of university or professional learning. In a review, Kebritchi and Hirumi (2008) revealed that many contemporary games claim to be designed in line with the ideas of social constructivism and situated cognition. Yet, for most of them, the authors didn't actually go on to validate the outcomes of learning with these games. A similar conclusion, along with an emphasis on the weak transferability between domains, has been drawn in (Hays, 2005). One notable difference between the research of Reeves and Malone (2007), and the rest of the discussed studies is that the first is exploratory, and thus identifies relevant individual evidence, whereas the others might try to capture learning for entire groups of students where individual differences introduce more interference.

There has been little discussion of the cost involved in developing a successful serious game. De Freitas and Jarvis (2006) stated that the focus of development must be on target users of the games, and recommend a series of user studies like semi-structured interviews and workshop activities. In another paper, de Freitas and Oliver (2006) put forward a framework to evaluate serious games in four dimensions. These are pedagogic considerations, learner specification, context and mode of representation. De Freitas and Oliver provided questions that outlined the particular aspects of their dimensions, that are shown in Table 3. This framework is employed in Chapter 3 to provide comparative analysis of existing games in project management and information security.

At a workshop in 2009 at the Game Developer Conference, three emerging trends for

Context	Learner Specification	Pedagogic Considerations	Mode of Representation
What is the context for learning? (e.g., school, university, home, a combination of several) Does the context affect learning? (e.g., level of resources, accessibility, technical support) How can links be made between context and practice?	Who is the learner? What is their background and learning history? What are the learning styles/preferences? Who is the learner group? How can the learner or learner group be best supported? In what ways are the groups working together (e.g., singly, partially in groups) and what collaborative approaches could support this?	Which pedagogic models and approaches are being used? Which pedagogic models and approaches might be the most effective? What are the curricula objectives? (list them) What are the learning outcomes? What are the learning activities? How can the learning activities and outcomes be achieved through existing games or simulations? How can the learning activities and outcomes be achieved through specially developed software (e.g., embedding into lesson plans)? How can briefing/debriefing be used to reinforce learning outcomes?	Which software tools or content would best support the learning activities? What level of fidelity needs to be used to support learning activities and outcomes? What level of immersion is needed to support learning outcomes? What level of realism is needed to achieve learning objectives? How can links be made between the world of the game/simulation and reflection upon learning?

Table 3: Dimensions of the educational games evaluation framework (de Freitas and Martin 2006)

serious games were identified. Among those were the broader use of early prototyping and new market demands, notably pressure for lowering costs. This trend can be seen as a response to the fact that commissioned projects which usually required multi-million-pound budgets and diverse teams working over years to deliver (Ulicsak and Wright, 2010).

Addressing this apparent difficulty in developing serious games, a number of researchers have explored different types of frameworks that could provide guidance towards the development of successful serious games. Winn's (2008) Design, Play, and Experience framework (DPE) and de Freitas and Oliver's (2006) evaluation framework are two that have already been considered there. Some others are considered in the rest of this section.

Nadolski and colleagues (2008) take a software engineering approach to define a methodology and a development toolkit (called *EMERGO*) for the design of serious games. Focusing on a broad coverage of learning, they do not even consider the aspect of enagement. Yet, for the purposes of learning, in their analysis phase, they put forward a list of high-level questions to define their requirements. Among those questions are ones that

address whether the games are being used as standalone learning experiences and the possibility that the game amounts for several academic credits. They also question whether the game is going to allow interaction between peers; how the game is going to be supported; what is the cost of use; and how will intellecutal property rights be managed. Whereas these issues might be relevant for a successful game, they sound premature for the limited interaction environment that is being presented in the publication.

Isbister, Flanagan and Hash (2010) also made recommendations about serious games design, but it is based entirely on interviews with game designers developing games for entertainment purposes. They summarised the importance of fun, final polish, learning of game content, integration of learning with engaging elements, collaboration and specialisation, role-play, exploration of ethical dilemmas and exploration of systems. No doubt fun is central to a game, even if it is meant to be serious. The three authors also suggest that polish could be achieved by adopting innovative approaches form experimental games – something that has been attempted here with the development of the *CCO toolkit*. However, when discussing learning of game content, the immediate feedback necessary for games might not correspond to the ambiguity introduced by complex learning content. Games might provide affordances to learn relationships that are inaccurate due to the limited realism. Often this is why mastering a game does not necessarily transfer to mastering related activities in the real-world. Another aspect of the same issue is the need to deeply link gameplay and learning, so that players cannot succeed in the game without mastering the intended learning objectives. This is why it is important that game scoring does not encourage surface learning. In a realistic scenario when players are asked to assume different roles (as in both *vLeader* and the *CCO toolkit*), the specialisation and role-play as discussed in the interviews overlap. As for issues of exploration, and whether it addresses moral and ethical dilemmas or systems, ideally this would depend on the category the learning objectives fall into. Isbister and colleagues' short paper provided guidelines, and it is left to game designers to find how these could contribute to the development of a serious game.

In his review of game design strategies (which can be considered as a higher level of game design patterns) Dickey (2005) considered different ways of engaging players into the game experience. He also discussed implications of these approaches to the instructional design that are necessary for a focused learning experience. Dickey, continued an already established trend by others, explored games as a successful media with the purpose to

adapt it for the design of educational materials. He considered several viewpoints that researchers before him had taken. On one hand he looked at viewpoints focusing on the potential of collaborative learning and peer support in the safe environment of multiplayer games, and on the other he brought up the motivational benefits of intrinsic factors such as challenge, controlled autonomy, discovery and fantasy. Dickey considered the importance of clear objectives and immediate feedback that reinforces engagement in games. He also compared all these to learning and the role of the teacher as facilitator and underlined similarities of conclusions of cognitive and constructivist research (compare to Sections 2.1.1 and 2.1.2). Dickey naturally related constructivism to social aspects, and cognitive approaches to factors of internal motivation and learning. While certainly such connections are very strong, in literature there is even broader overlap. On one hand, quite distinct from social constructivism, there is the concept of intrinsic constructivism, where an individual develops their own meaning, possibly opposing various social influences. This can hardly be attributed to the immediate social environment. On the other hand, contributing factors like worked examples, feedback, challenge are more genuine and personal when they apparently involve another human being.

In his framework Dickey considered player perspective, narrative and interaction as game design factors and what implications they could have on instructional design. As seen from Dickey's discussion, the perspective a player takes undoubtedly affects their view of both task and environment, it is often strongly influenced by the nature of activities that learners need to be engaged in. For example, conversation and negotiation (as in the case of Study 1 – the proxy study with *vLeader*) might suggest first-person perspective, whereas the implementation of interventions (as is the case of the *CCO toolkit* developed here) might suggest a broader – thus probably isometric – perspective.

When considering narrative in games Dickey focused on *backstory* and *cut scenes*. Despite his dual perspective of intrinsic and social drives to learn, and his reference to character-based narratives, Dickey chose to separate narrative from roles and characters. Yet, when considering implications for instructional design he discussed both of the explanatory power of narrative and role-play, relating both to case studies. As with a many serious games, the backstory of *vLeader* and the *CCO toolkit* is set in a realistic setting. Thus, the broader setting is implicit and not delivered. The parts that are actually delivered, merge with the key cut scenes of the games that serve as introduction to the separate levels.

Dickey's view of interactive design also considered settings, as in spatial, temporal,

environmental, emotional and ethical dimensions, which again, similar to narrative are trivialised in realistic game designs, such as the ones considered in this thesis. Dickey only superficially touched on the complex decisions and broad choices of actions, feedback and affordances.

Gunter, Kenny and Vick (2006) also suggested that a formal design paradigm is necessary for serious games. They considered three theories from instructional design and discussed how these could be applied to serious game design. The theories they considered were Gagne's Events of instruction (Gagne, 1985), Keller's ARCS model (Keller and Kopp, 1987) and Bloom's Taxonomy of the cognitive domain (Anderson et al., 2000). Gunter and her colleagues matched elements of Gagne's and Keller's models to common game elements. However, their mapping remains unconvincing in that it underlined the importance of engagement, but did not suggest means of achieving it. One aspect of this shortcoming is that they provided suggestions, but did not ground them neither in examples of existing games, nor in a design of their own.

Yusoff and colleagues (2009) put forward a short position paper that proposed a conceptual model for serious games. Their model is represented by a structural diagram that combined learning content with game elements. They proposed 12 game attributes. There are obvious overlaps between some of these. One such overlap is scaffolding being mentioned in parallel to other attributes, like incremental learning on one hand, and practice and drill as a way to increase difficulty on the other. Another overlap suggested rewards as something distinct to providing feedback. The authors did not explain whether they intended to address different goals with different attributes. One such possible distinction could be incremental learning and rewards addressing motivation, as opposed to practice, along with feedback addressing learning.

These frameworks are insightful attempts to provide a structure to the development of serious games. They explore how different components contribute to the experience of serious games, thus becoming useful analytical tools to demonstrate how different design decisions can lead to different learning experiences. However, similar to the works of Gunter's and Isbister's teams, the work of Dickey and Yusoff stop short of suggesting how the sets of their heuristics can be integrated towards a successful serious game. They do not address the key challenge in designing serious games – how to combine engagement and learning techniques in a way that results in coherent experiences which focuses on the exploration of the intended learning content.

In the design of learning experiences de Freitas and Oliver (2006) focused on the settings in which games are used. For this reason their framework is a useful complement to the design of serious games and digital instructional materials in general.

2.5 Design-Based Research

This research started as pure evaluation of serious games in learning and engagement, due to external circumstances (see start of Section 7.2) it had to evolve into a development project. This naturally led to a method featuring both development and research, which can be thought to fall into the wider notion of *design-based research* (DBR) as it emerged in the early 2000s (Anderson and Shattuck, 2012). Design-based research emerged in the tradition of developing practical interventions in educational research. It has also been called *development research* (van den Akker, 1999) or *developmental research* (McKenney and van den Akker, 2005). These alternative terms reflect more closely the perspective taken in this thesis, as in this particular case the intervention design is represented by the development of game-based learning tools. It is similar to approaches like the one of the *reflective practitioner* (Schön, 1991) and *action research* (Reason and Bradbury-Huan, 2001) – aiming to contribute to both theory and practice.

Features that shape design-based research as a valid qualitative and participatory research method were outlined by Anderson and Shattuck (2012). They focused on being situated in real educational research, focus on the design and evaluation of an intervention, the use of mixed methods, the use of multiple iterations, the required collaboration between researchers and practitioners, and finally, the interaction between the process and design principles (Anderson and Shattuck, 2012).

The adoption of DBR for the development of serious games had been suggested by Squire (2007) among others, but there is hardly any documented adoption of this idea, a notable exception is the work of Dunbar and colleagues (2013). However, the absence of literature on using DBR as an approach to develop serious games is not a rejection, but rather a symptom of similar methods established under different names, that are already widely used in serious games development. An illustration of this is the fact that Squire is a co-author in Dunbar's publication, yet DBR (phrased as "design research") is mentioned only once in the paper.

However, there is a range of literature that uses approaches similar to DBR for the development of serious games. These methods commonly examine the perceptions of the

real-world users and domain experts using adopted unstructured methods for data collection. It is also typical for them to extract recommendations for the development of a serious game, and employ multiple interations. Such examples can be found in user-centred research (Vasalou et al., 2012), (Karpouzis et al., 2013), participatory simulations (Kreitmayer et al., 2012) and semi-structured requirements elicitation (Seager et al., 2010; Swanson and Jhala, 2012).

In the next section the aggregated principles from the literature review are outlined. They are further discussed in Chapter 4 in the context of the development undertaken as part of this thesis. The principles are refined throughout the studies conducted and reconsidered in Chapter 10, as an iteration in "advancing credible assertions" (Barab and Squire, 2004).

2.5.1 Derived Principles

In line with DBR, a set of principles were derived from this literature review which were used as a starting point for the development of a game-based intervention. Five principles can be derived from the review conducted in this chapter:

1. learning needs to be guided pedagogically;
2. the learning environment can support engagement to learn;
3. learning itself needs to be incrementally supported (in other words, scaffolded);
4. learning unfolds better when done within learning communities;
5. an adaptive learning experience can potentially develop various forms of learning.

Chapter 4 contains further elaboration on how each of these influences both methodologies – for development and evaluation. Davies and Mangan's (2006) process from theory through pedagogical principles to learning activities provides a broad framework of how learning needs to be delivered. The method of implementing this is further instantiated by Marton's (2004) perspectives towards the *object of learning*, and is to some extent similar Laurillard's (2001) *conversational framework*. These perspectives represent the main pedagogical principles of iterative teaching and probing the learners' understanding for how the theory has been perceived.

Within each iteration learning needs to happen incrementally. This represents appreciation for the fact that the intended *cognitive schema* that is meant to be taught, is probably too complex to be understood in its entirety at once. This is why learning needs to happen gradually. More specific instructions for how this should happen were provided by

Marton and Pang (2006) with their necessary conditions of learning. This was also advocated by researchers in cognitive load theory (van Merriënboer and Sweller, 2005) where it is termed *whole-task sequencing*. One very specific tool to achieve the early stages of this incremental learning are *worked examples* (Sweller, 2006). A key feature in traditional learning that has fallen behind in early technological solutions is learning in a community. Research on *communities of practice*, *role-models* and *peer learning* has shown how important these have been in more complex and subjective learning domains. Learning in a community allows for discussion and feedback, and more generally for *inquiry learning*.

Yee's research on motivation to play (2005) has shown the beneficial effects of *gamification* (as a way to introduce quantitative feedback and competition) and social interaction within the virtual environment. Even though Yee's research was descriptive (it represents what users find in already existing platforms), it also has the potential to be used perscriptively (as guidelines for what to be included in a platform to make it successful). When considered in the context of the other principles listed here, it becomes apparent that both social interaction and quantitative feedback have pedagogical benefits as well.

Finally, the complexity of the learning activities within a serious game suggests that different forms of learning might be happening at the same time. Due to the different possible forms of learning, a corresponding portfolio of assessment techniques needs to be employed as a way to capture various forms of learning.

The discussion of how these principles can be applied and integrated into a methodology is continued in Chapter 4 after the review of currently existing technologies in the next chapter.

Chapter 3: Review of the State of the Art in Serious Games

Current serious games in the domains of project management and information security are reviewed here. The former of these two has been prominent for decades and there is a broad range of attempts to develop interactive technologies for learning. The latter is relatively young and has grown recently. Correspondingly, attempts to develop serious games for it have only been made in the last 5 to 10 years. Due to the broad variety, this review discusses game examples that I considered representative, with the full list of considered game titles attached in Appendix A. The serious games are compared in the evaluation framework of de Freitas and Oliver (2006). It evaluated serious games according to four dimensions: pedagogic considerations, learner specification, context and mode of representation. These are discussed in detail here.

For *pedagogic considerations*, they ask questions about pedagogic models and learning outcomes. As summarised by Kebritchi and Hirumi (2008), educational games commonly do not have extensive theoretical underpinning. Some noteworthy exceptions have been discussed in Section 2.3. In the common situation that games are being embedded in a learning module (like *CyberCIEGE* and the majority of games on project management) they need to somehow relate to the corresponding class objectives. In such situations the learning objectives are derived from the class objectives. The study by Kebritchi and Hirumi (2008) revealed that commonly these objectives do not drive the learning activities within the resulting serious game in a pedagogically informed manner. Yet, it is typical for experiential learning with games that it involves role-play, perspective taking, problem-solving. Some games provide more immersive role-play (*Sharkworld*, *Race-to-Results*, *CyberSIEGE* and *VOLT* all take different approaches to do that), or embed play from different perspectives (like *vLeader*, where the player's role in the game evolves to allow for this). Whenever games are multiplayer, they provide affordances for learning discussion and social learning, adding to the learning practices that players engage in. Such reflective discussion is one important factor to learning that is easier to stage in classroom environments (again as the majority of considered games here). Another is briefing and debriefing sessions, respectively before and after the use of a serious game. These are also a common practice when there is focus on complex knowledge, and conditional knowledge in particular. Whenever the learning involves interpersonal or social interaction, there are no clear-cut correct answers and debriefing is beneficial, as explained by Fanning and Gaba (2007), among others.

De Freitas and Oliver dedicated their second dimension to the *learner specification*. Here they ask questions about the learner's background and preferences. Whenever learners form groups, this dimension explores collaborative approaches and possible means of learning support. Typically, serious games are designed for a very wide range of audiences. No matter if it is meant to be played at home, or in a class, the serious game could reach very different audiences with various backgrounds and preferences. Even when participative design is being used, it involves learners that can only be a proxy for the much broader audiences that subsequently use the resulting serious game. This imposes limits on designers to assume as little as general interest to the field that drove to the learner playing the game. In class settings, specialist background knowledge is typically expected. Yet commonly supplementary facilitation or materials are provided to help learners better grasp the learning experience with the serious game (the supplementary materials used for *vLeader* are described in Section 5.2.2). Learner preferences are often channelled into customisation (like in *TOPSIM* or *CyberCIEGE*), range of game goals (see Figure 8 in Chapter 5 for an example how this was done in *vLeader*), group forming among other techniques. In class settings games are commonly played in teams (*TOPSIM*, *Race-to-Results*, as well as the majority of games played in class), thus, adding teamwork, competitiveness, discussion at different levels and other social learning activities to the learning experience.

The third dimension, *context*, explores the learning environment. Beyond considering whether games are targeted at classes or individuals, it also looks at how this context affects learning and how the context can be linked to practice. The two typical use cases of serious games are formal use in classes, and informally in private at the learner's own pace. Typically in classes learners have common background, access to facilitation and shared experiences, so there is access to feedback and discussion. The context can contribute to future applications to practice by invoking links between the game environment and the real world.

The final and fourth dimension is the *mode of representation*. De Freitas and Oliver explored how well the serious game could support learning activities, including fidelity, immersiveness, realism. Here also ways are sought to make the knowledge transferable to practice. Computer games provide many new affordances, simulated environments (*TOPSIM*) and virtual worlds (*CyberCIEGE*, *vLeader*, *VOLT*) being two of them. Serious games often immerse players in 3D virtual settings, but there is a tradition in simulations

for non-animated (like *Race-to-results* and *Sharkworld* where video sequences are used), and are commonly less graphics-intensive applications (*Hardvard's game*, *TOPSIM*). There is also a long-standing tradition of paper- or card-based games, exemplified here by the card games on information security awareness reviewed at the end of the chapter. Serious games can be more relevant to practice or not, depending on the choice of data (see *TOPSIM*), narrative (like in *Sharkworld*) or appearance (like in *CyberCIEGE* or *VOLT*) and how these relate to practice.

A discussion of serious games and related virtual environments in the domains of project management and information security follows.

3.1 Project Management

As reasoned in Section 2.2, four main streams are considered to be of interest. These are the traditional three perspectives towards games: games, simulations and interactive storytelling, along with the slightly more different perspective of virtual worlds.

By far, the most widely adopted approach is the one of simulation games – environments that inherit properties of simulations (like the predominantly realistic game environment) and games (for example competition and user-friendliness). The distinction between games and simulations was discussed in Section 2.2 and more broadly by Sauve and colleagues (2007). The development of simulation games about project management has followed the wider trend of development in educational games. A number of courses based on simulation games, which have gained recognition by for example the Project Management Institute (PMI), delivering the most widespread project management certification procedures. One crucial feature of the branch is the two rather poorly-connected streams of development: one of the scholarly development processes and publications; the other – of market- and revenue-driven industry development.

The applications reviewed here encompass examples from both of these streams. The criterion for inclusion was that entries should have received some form of coverage online. The method used here was to search for "project management" in combination with "game", "simulation", "virtual world", "second life", "digital storytelling", "interactive storytelling" successively. Then the first 50 entries on both Google Web Search and Google Scholar were reviewed. Scientific publications from 1996 on have been investigated in detail. For a discussion of earlier developments, see (Elgood, 1997).

The identified products were categorised according to their self-description. The majority

of them tend to be described as simulations. One possible reason for this might be a marketing perception bias against games. The investigation resulted in 2 games, 17 simulations, 2 interactive storytelling applications, 3 virtual worlds, where games that have commonalities (like the cases, when two games are variations of the same core development) are considered as one. There were 14 more entries that for various reasons did not meet the predefined conditions, for example when they did not have sufficient documentation (*StevensTech Capstone Simulation*, *VOLT*); did not involve computer support; were older than 1996 and no current information was available; or were in some development phase (*Adam Montgomery's game*, *Vivian Valiant*). Exemplars are discussed here and compared in Table 4. The full list of reviewed games can be found in Appendix A.

All reviewed games develop a specific story to represent the learning context, usually putting the learners in a project manager's position. Some (like *Sharkworld*, *SimulTrans*) feature an already started project as part of the story.

There seem to be several categories within games and simulations employing similar

	Pedagogic Considerations	**Learner Specification**	**Context**	**Mode of Representation**
Sharkworld, SimulTrans	Project Management	non-professional users	informal, online	individual play in gamified website
TOPSIM, Leadership-in-Action, Harvard's game, CBT module, Contract and Construct, Simulation laboratory, AXL	Project Management	class attendees	formal in class	individual play in simulation game
Synergest's game, G2G3 Polestar, ManSims, Sim Project, Family Life, Race-to-Results,	Project Management	corporate staff	formal in class	computer-assisted business simulation for teams
vLeader	Leadership and Negotiations	class attendees	formal in class, self-paced at home	individual play in 3D environment
4D virtual construction	Project Management in construction	class attendees	formal in class, self-paced at home	individual play in 3D environment
VOLT	Military Leadership	military officers	formal, in immersive environment	immersive 3D environment
Telespace, MPK20	Project Management	corporate staff	formal at workplace	collaborative 3D environment

Table 4: Comparison of serious games and related applications on the domain of project management. Based on the educational games evaluation framework (de Freitas and Martin 2006)

patterns. The group of games that seem to have attracted most commercial attention is a variation of traditional table-top games which make use of a common computer-supported dashboard (like *Synergest's game*, *G2G3 Polestar*, *ManSims*, *Sim Project*). Variations make use of excel sheets to make calculations easier (*Family Life*), interactive dashboards to share data and even pre-recorded video sequences to enrich the story of the experience (*Race-to-Results*). These games are essentially workshop-shaped group learning activities, led by a facilitator that manages both team dynamics and learning content. An advantage of these games is the fact that free conversations, beyond the restrictions of the digital environment, are possible between learners or with the facilitator.

Another group of simulations takes a more technical approach. These project management simulators put learners in one single role, although many of them are intended to be played by small teams. A long-improved classic is *TOPSIM*, deployed around the world for several decades. Each temporary deployment is customised for the particular context. A number of these project management simulations are closer to computer-based role-playing games within a simplified virtual environment (for example, *Leadership-in-Action*, *Harvard's game*, *CBT module*, *Contract and Construct*, *Simulation laboratory*). Usually these also have a more expressed game story and some form of virtual characters (*Sharkworld*, *Prendo*'s games, *SimulTrans*). Communication with virtual characters is based on dialogues with predefined choice of options. A somewhat different approach was taken in *vLeader*, where interaction is done at the level of speech acts in a conversation (Aldrich, 2003). This allowed the learners to focus on the *illocutionary force* (meaning the intention behind what is being done) of interactions, thus, enabling complexity of interaction, while still maintaining relatively simple vocabulary and repertoire.

For the purposes of management in construction, a technology called *4D virtual construction has* been developed, that allows for an actual virtual construction process. It adopts a 3D perspective and tracks the time as a fourth dimension (hence 4D). For the purposes of this review, it is considered alongside applications that simulate product development and focus on the more technical side of project management.

The two storytelling applications for training have been developed by the Institute for Creative Technologies at the University of Southern California – one of the leading institutions in interactive storytelling worldwide. They were both developed for military training purposes and are specific to the type of field endeavours, which military leaders need to engage in. One of them, *Army Excellence in Leadership* (AXL), adopts an

interactive case-method teaching that employs pre-recorded video sequences. In its video-based approach that provides a much more specific context, AXL is similar to HP's *Race-to-Results* and *Sharkworld*. The other one currently being developed is called *Virtual Officer Leadership Trainer* (VOLT). It is being delivered within a training room environment. Through interaction with virtual characters, it is intended to deliver practice of communication skills to complement traditional learning methods.

Finally, three virtual worlds have been identified to be reviewed here – *SecondLife*, *Telespace* and Sun's *MPK20*. In their specialisation, these three cover the range from general purpose through customisable on demand to dedicated development. All these have been adopted for some form of learning, including in the domain of project management. However, in contrast to the previous categories, this software was not developed for the purposes of project management training in particular, but rather, only to serve as general infrastructure to facilitators that conduct project management training sessions. As a consequence, the use of virtual worlds for training is very diverse. In line with this, various authors focused on the identification of possible uses of the platform for educational purposes. For examples of this see (Ryan, 2008). Thus, virtual worlds actually afford themselves as environments for learning where multiple activities are possible, as opposed to learning tools, which are matched to a particular activity.

3.2 Information Security

Fogg's (2009, 2002) captology has been used by researchers at Carleton University to design a password creation system – an example of usable security tools aimed at non-security staff (Forget et al., 2007). Beyond such endeavours, which are actual productivity tools, and not technologies for learning, there are not as many games about information security, as there are for project management. Due to this sparsity a review of interactive technologies for learning in the wider domain of information security needs to take a broader approach. As a consequence, non-digital cases are also considered, and these are commonly card games. The 4 games considered are displayed in Table 5 and discussed below.

A number of existing serious games for information security already exist. For example, the Naval Postgraduate School developed a game to spread awareness about cyber security called *CyberCIEGE* (Cone et al., 2007; Irvine and Thompson, 2004). It is a customisable platform that allows designers to develop scenarios for their organisations. A typical scenario in *CyberCIEGE* is about preventing users from letting malware into the corporate

	Pedagogic Considerations	Learner Specification	Context	Mode of Representation
CyberCIEGE	Introduction to Information Security	class attendees	Formal in class, self-paced at home	Individual play in 3D environment
Elevation of Privilege,	awareness about cyber threats	groups of non-professional users	Informal, presence	Group play with paper cards
Ctrl+Alt+Hack	awareness about cyber threats	groups of non-professional users	Informal, presence	Group play with paper cards
Privacy game	awareness about privacy	groups of non-professional users	Informal, presence	Group play with paper cards

Table 5: Comparison of serious games and related applications on the domain of information security. Based on the educational games evaluation framework (de Freitas and Martin 2006)

intranet, preventing social engineering and safeguarding data. Two small-scale evaluation studies were conducted to compare learning with the game, to learning with an educational video (Jones et al., 2010). Results from that evaluation indicated slightly better (no statistical significance reported) results of the game-playing group. Better performance was illustrated with more elaborate answers, and authors linked it to the fact that students that used the game spent much more time with the corresponding learning materials.

Another example of a serious game to teach in the domain of information security had been presented at the SOUPS 2013 conference. Unfortunately, little is publicly available about it. Of this little information, probably most insightful is a comment made by Ross Anderson on his blog (Anderson, 2013): “Eva Vuksani finally talked of *Device Dash*, a game in which a system administrator tries to keep a corporate network free of compromise. There are users attaching bad devices, and administrator compromises that spread to all nearby users and devices; for defence there are scanners and network access control. The game is more like space invaders or tetris though than a strategy roleplay.” My attempts to get in touch with the Vuksani regarding any further information or studies proved unsuccessful.

This description is also apt for three developments of card games in the domain of information security: *Elevation of Privilege* (Shostack, 2012), *Ctrl-Alt-Hack* (Denning et al., 2012) and the *Privacy Game* (Barnard-Wills and Ashenden, 2013). All these take an existing popular *game mechanic* and develop an information security narrative around it. This way they make sure to preserve the game’s attractiveness, but fall into the trap of delivering games that are possible to play and win without engaging in the information

security narrative, and thus, not learning at all. This is possible because the scoring system is not inherently linked to some form of learning assessment aligned with transparent intended learning objectives. In light of Yee's motivations to play online games (2005) it could be speculated that the narrative could help learners that like to immerse, but not those who are driven by achievement.

3.3 Discussion

As dealing with people is a major part of project management and information security, it is featured in some form, in all reviewed virtual environments or games. In some of them, this social element is narrated (for example through the use of virtual characters like in *Sharkworld*) or simulated (*vLeader*).

Yet, there are an array of multiplayer games where social components are represented by interaction between players. This reduces the need to involve professional narration or *virtual characters* in the development. Thus, making development more accessible to a small team, or a single developer (as was mostly the case in this thesis).

Instead, it requires the environment to sufficiently engage other players to assume the antagonist role. In the context of class activities, such an environment is extended by facilitation (like *Race-to-Results*, *TOPSIM*).

However, a practical consideration when doing research is that typically the majority of multiplayer games are synchronous. This means that players need to play simultaneously, and because of this their individual contributions are difficult to isolate for the purposes of controlled studies. Due to this asynchronous models for interaction are of interest, so that peer contributions can be fixed and interaction of individual users can be subjected to experimentation.

Chapter 4: Methodology

4.1 Review of Methods Used

Design-based research (DBR) has emerged as a commonly accepted approach to the design of educational interventions (Anderson and Shattuck, 2012; Barab and Squire, 2004). Authors in the field emphasised that while traditional educational research is focused on demonstrating improvement in learning, design-based research rather focuses on the process to achieve these improvements (Herrington et al., 2007). Features that shape design-based research as a valid qualitative and participatory research method are outlined by Anderson and Shattuck (2012). They focused on being situated in real educational research, focus on the design and evaluation of an intervention, the use of mixed methods, the use of multiple iterations, the required collaboration between researchers and practitioners, and finally, the interaction between the process and design principles (Anderson and Shattuck, 2012). An explanation for the approach towards these features, taken in this thesis follows.

Ultimately, this piece of research aims to contribute towards the use of serious games in a real-world setting. This is one of the features typical for design-based research and has to do with the contextual specifics in real-world applications. These specifics very often make it difficult to transfer results of lab-based experiments to a less-controlled context, as exemplified by the work done in Chapters 9 and 10.

Another typical characteristic of design-based research that can be observed in this work is the use of mixed methods and measures. Common for the studies done within this thesis is that they involve both quantitative and qualitative assessment of learning and engagement, interviewing learners, and analysing data collected by the used game. I attempted to cross-validate (a process also referred to as *triangulation*) the results of these methods whenever this was possible.

Design-based research is commonly conducted in classes – a setting that typically gets repeated annually. This allows for the research effort to be continuously refined in multiple iterations in contexts that are similar over the years. Such an iterative approach in class is referred to as *transactional inquiry* by researchers in the field of *threshold concepts* (Cousin, 2008). Due to the expected duration of PhD research, this was not considered to be viable. Instead, an iterative approach, where class studies are alternated with lab studies, in order to allow for higher frequency of iterations.

A characteristic feature of design-based research is the necessity for intensive and long-term collaboration between researchers and practitioners (Herrington et al., 2007). Anderson and Shattuck (2012) refered to the necessity of involving teaching practitioners to work with educational researchers. Yet, authors from other fields suggested involvement of other types of practitioners, for example in their work on *threshold concepts* Meyer and Land (2006b, p. 4, 2003) discussed the involvement of practitioners from the studied domains. All this comes on top of the complexities related to the development of a game, including software development, artistic design and interaction modelling (game design). Herrington and colleagues noted one advantage that students doing design-based research could possibly have, when their supervisors have experience with this research approach. This was not the case with my work, but on a more general level there was a very high appreciation of collaborative real-world research.

I conducted the early part of my research within a large-scale collaboration with the TARGET project (Andersen et al., 2009; Fradinho et al., 2011), but there were coordination challenges that undermined the project's overall results and reduced the possible scope of my contribution. In that situation I found a way forward in developing a prototype for a serious game that employed a practical research framework for crime science developed by one of my supervisors (Ekblom, 2011a, 2001), and adapts it to information security. This was done in close collaboration with the author and experts from the targeted domain of information security. This collaboration was more successful in delivering a tangible result (in particular work done for Chapter 6 to Chapter 8).

The final key characteristic for design-based research is the interaction between the process and design principles. It is the one that makes the method a valid form of research, in that it combines theory – the employed and accumulated principles – with empirical evidence – the real-world data gathered during formative studies. My starting principles are a collection of recommendations from the studied domains, educational research, software and game development and usability as these were reviewed in Section 2.5. My conclusions, with their limited generalisability, are summarised in Section 7.3.

In their guidelines on how to organise design-based research within a PhD thesis (see Figure 6), (Herrington et al., 2007) refer to a diagram by (Reeves, 2006) in which four distinct stages of the research process are identified. These are i) analysis of the practical problem, ii) development of a solution, iii) iterative refinement cycles, and iv) reflection of the form of design principles.

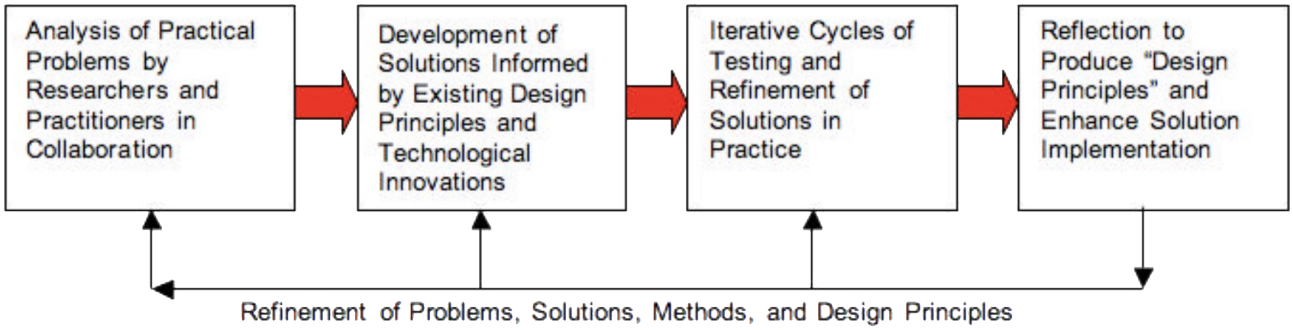


Figure 6: The design-based research process, as described by (Reeves 2006) and referred by (Herrington et al., 2007)

I identified the need to customise the learning tool previous studies (Ruskov et al., 2010; Seager et al., 2010) and validated it as part of the study reported in Chapter 5. As a result, the *CCO toolkit* was developed – a prototype for a serious game. DBR was employed to study conditional knowledge developed in the context of working with people in information security. The toolkit was designed to help learners address the necessary complexity of the recently developed CCO framework (Ekblom, 2011a, 2001). The development is reported in Chapter 6.

The need for this project grew from the practical constraints related to teaching the crime prevention framework on one hand, and the need for a similar approach in information security on the other. Teaching the crime prevention framework has shown evidence that potential users find its complexity overly challenging, notably because other more established models, like the *crime analysis triangle* (Clarke and Eck, 2009), are simpler. In the domain of information security attempts have been made to bring together technical and personal factors, like (Blackwell, 2012), but not social ones. This is one of the reasons it was suggested that CCO is adapted to information security (Collins and Mansell, 2004).

This led me to try to develop a software toolkit (with the intention that it grows into a serious game) that guides users to use the CCO framework. This had to include an adaptation of the framework to information security. This adaptation is a research project on its own, and is not included it in this thesis. Yet the conducted research delivered as a by-product recommendations of how CCO could be modified to better match the needs of information security. This was a consequence of the constraint that the project was not part of a wider team effort.

In this chapter two different methodological reviews are made. One of them addresses the approach towards development and considers methodologies used in the development of *serious games* and related products, including learning experiences, games and persuasive

technology. The other review addresses evaluation methodologies again from several research perspectives: usability, learning and expertise, and finally games and engagement. The reason for this separation is the range of communities that have contributed to the development of serious games and their corresponding differences in focus.

4.1.1 Development

The two development areas that are of particular relevance to serious game design are learning design and game design. The focus in the former is designing experiences that provide opportunities for efficient awareness and learning. The latter designs focus more on motivation and engagement of their users, in this particular case – learners.

Learning Design

This review here considers state-of-the-art learning design, thus, practices from social constructivism, development of *personal learning environments* (PLE) and theories around threshold concepts and variation are discussed.

Ideas around social constructivism have become widely appreciated among teaching practitioners. Notably the idea that knowledge is not simply transferred from the knowledgeable to the ignorant, but rather that novices construct their own understanding of new concepts supported by facilitation. Knowledge construction is actually its cognitive integration with previous experiences. From this perspective cognitive scaffolding; forms of support; and differences between the perspectives of teacher and learner are considered.

It has been widely agreed and verified that in such a process of help in the form of guidance (or more formally *cognitive scaffolding*) increases the potential and effectiveness of learning. In this thesis scaffolding has been discussed in the context of PBL and IL (see Section 2.2.2). Cognitive scaffolding has gained popularity and evidential proof recently, but it has been used in an extremely wide variety of contexts, diluting the concept as a result. This ambiguity had been exemplified in greater detail by Pea (2004). Kollar, Fischer and Hesse (2006) distinguished two types of scaffolding with respect to collaborative learning: scaffolds that provide support on a content-related or conceptual level and scaffolds that provide support related to the interactive processes between the collaborators – also called collaboration scripts. Both of these are of interest for this research effort, albeit the latter appears to be more approachable by technological means.

In line with the ideas of *situated cognition* and *role models*, various cognitive scaffolding

techniques have been used in working practices. Regardless of the wide variance between specific instantiations or whether these are called tutoring, coaching or mentoring, such techniques are being successfully adopted all over the world. One way to distinguish between them could be the argument that *coaching* and *tutoring* have the purpose of helping the learner to better understand the subject and learn new knowledge. *Mentoring* has the purpose of helping one to better understand the environment and support one's decision-making (compare with Kollar' findings). Another alleged difference is that tutoring involves assessment, whereas *coaching* and *mentoring* commonly rely on the (possibly indirect) assessment of workplace performance.

Devising a portfolio of assessments is work that has been considered as part of curriculum and assessment design. Biggs (2003) for example, discussed of the two different perspectives that teachers and learners take, and in the eyes of the pragmatic learners, learning starts with considering assessment and the objectives that it is going to measure at the end. These views have made a robust case that planning for assessment needs to start with a plan for the learning activity.

One of the developments in the social constructivist tradition is the study of *personal learning environments* (PLE). Researchers in this field examined and reinforced emergent practices in using online technologies to support learning. Attwell (2007) sees in the PLE concept the potential for bringing together learning and knowledge creation (for example for research and development) in organisations. Kompen, Edirisingha, & Monguet (2009) outline some benefits of PLEs for the individual learner: the ability to organise and manage information that is already known, as well as the access to new sources; the chance to integrate tools they have been using with ones that are new and useful to them; the use of the PLE to filter out relevant information. There is evidence that some learners associate entertainment and socialisation with personal learning environments (Kompen et al., 2009).

A wide variety of Web 2.0 tools can be part of the mash-up that typically comprises a PLE, as this was indicated, for example by Trinder, Guiller and colleagues (2008). In an examination of such tools, Kompen and colleagues (2009) outlined four distinct types of PLE, each defined in terms of the tools that are most heavily used: wiki-based, social network-based, aggregator page-based and browser-based. These categories are not exhaustive, but indicate preferred patterns of use. In relation to this, Kompen and colleagues recommended a bottom-up approach towards building a PLE, meaning that

they would leave the initiative to individual learners to select and bring together the tools that they prefer to use. From this perspective, serious games are only one of the tools that can be blended together with the mash-up that forms a PLE. Yet, one could investigate how these main patterns of use match similar patterns of use in gameplay.

Although distinct from the perspectives of teachers and learners, the scope of learning content is also an important decision that needs to be made during the development process. This is even more so when considering the development of serious games, as their development is more costly than traditional lecture preparation. The high cost of game development requires a very careful consideration of learning objectives and methods while designing a serious game. Research around *threshold concepts* tries to identify learning content that is more *troublesome*, but is *transformative* – its grasping allows for deeper understanding of the whole area. Threshold concepts are also said to be *irreversible* – once learned it is difficult to ignore them when thinking; *integrative* – enable the understanding of underlying interdependencies; and *bounded* – indicate boundaries of different subjects.

As Davies explicated, even the mere identification of *threshold concepts* is not a straightforward task (Davies, 2003) and there have been suggestions that these could possibly be culturally dependent (Hulpuş et al., 2010). So it should come as no surprise that the integration of threshold concepts in learning practice is also challenging. Davies and Mangan (2006) suggested this integration should undergo a process from theory to pedagogical principles to learning activities. The aim pursued in this research effort, is that these learning activities converge around *game-based learning*. Still, the open question that needs an answer is what pedagogical principles would connect the theory surrounding threshold concepts to game-based learning, and how could this happen. This is also aligned with the focus on practical principles that design-based research puts. According to Reeves (1999) as quoted by Herrington and colleagues (2007) "the purpose of [educational research] should be to improve, not to prove" interventions.

Although, the approach to each particular threshold concept might need to be unique and possibly also individualistic to the learner, there is one pedagogical principle that has been repeatedly brought up by researchers – the one of learning variation (Davies, 2003; Davies and Mangan, 2006; Meyer et al., 2008). Variation has been used in learning for many years, probably starting with the work of Dienes (1973), but as already mentioned Marton and Pang (2006) had developed it into to a systematic framework. Researchers, developing

the theory around threshold concepts and troublesome knowledge have argued that variation is exactly the pedagogical principle that would enable learners to grasp the complexity of a particular threshold concept. Possible evidence to support this could be sought in the parallel interests of both the *phenomenographic* (Pang and Marton, 2003), and threshold concepts (Davies, 2003) communities of the economics notion of supply and demand. The first group observed the effects of learning with the help of variation, the second identified widespread difficulty among learners in grasping the theory. Both focused on the parallel dynamics and interplay of the supply and demand curves as something particularly troublesome to grasp. Such a match has been reinforced by collaborative work, such as the one of Pang and Meyer (2010). However, it remains to be demonstrated that these findings in economics can be transferred also to other cases of troublesome knowledge in other domains.

Finally, findings of the phenomenographic community have converged on a theory of learning variation. Phenomenographics studies have revealed that influencing the learner's attention through variations helped develop comprehension of challenging phenomena. Variation theory focused on an object of learning, examined in a *learning study* and through the provision of the necessary conditions.

According to this theory a phenomenon may be understood in different ways, where each way of understanding is associated with a different subset of features relating to the phenomenon. When a certain phenomenon needs to be understood by students, variation theory considers it to be the object of learning (Marton et al., 2004, pp. 3–4), also considered by Winn (2008). Researchers and teaching practitioners are considering different perspectives on it, which take into account all three of the following: the student's understanding of the phenomenon (Marton's *lived object of learning*, or Winn's *experience*); the planned learning activities that would support students (Marton's *intended object of learning*, or Winn's *design*); and the actual activities as they happen (Marton's *enacted object of learning*, or Winn's *play*). Laurrilard's conversational framework (2001) illustrated in Figure 4 considers the conceptions of the teacher and the learner, as they are conceived (respectively *teacher's concepts* and *student's concepts*), and how they are enacted (respectively *teacher's constructed environment* and *learner's specific actions*). Marton and Winn's perspectives on one hand, and Laurillard's on the other complement each other in a way that they are not derived from one another. Thus, it seems that the objects of learning as described by Marton are the aggregate representation

of the interactions that happen between the objects of Laurillard's framework. More specifically, the teacher's collection of reflective feedback from the constructed learning environment and corresponding adaptations represent the alterations in the intended object of learning. Similarly, the interaction between this environment and the student's actions represent the enacted object of learning. Finally, the way the student reflects on their actions and the adaptations that they develop according to their understanding of the theory is the lived object of learning. All these considerations shed light on the idea that these objects are distinctive and need to be monitored when developing, and subsequently deploying a learning activity (and the corresponding learning technology). This thesis explores approaches to each of the manifestations of the object of learning, and thus indirectly also to Laurillard's interactions.

Marton's theory has resulted in a pedagogical framework, called *learning study*, which puts emphasis on the two notions of variation and invariance. The former essentially brings the concepts into focal awareness, while the latter draws the learner's attention away from concepts that are less relevant to that particular learning experience (Marton and Booth, 1997). Marton and Booth explained this retraction of the invariant elements in terms of either transcendence or a situation of being taken for granted, which helps in tackling that, which has been considered relevant.

Marton and Pang used the learning study framework for curriculum design (Marton and Pang, 2006). This process can be combined with a consideration of the theory around threshold concepts and troublesome knowledge. It seems reasonable to focus closely on the implications of threshold concepts for variation and allow learners to develop better insights. Furthermore, the knowledge that the particular learners bring with them to the learning activity needs to be investigated as a possible source of features to be kept invariant (Marton et al., 2004). A resemblance between variation theory and *cognitive load theory* could be drawn, in that both theories put emphasis on learner's attention and ways to guide it in incremental steps towards the comprehension of the particular topic being taught.

As part of their work on the variation framework Marton and Pang (2006) identified four conditions that they claimed to be necessary for learning. These conditions are:

1. *Contrast* – A certain quality cannot be discerned without the simultaneous experience of another quality that is mutually exclusive to it.

2. *Separation* – A dimension of variation, which can take on different values, cannot be discerned without other dimensions of variation being invariant or varying at a different rate.
3. *Generalisation* – A certain value in one of the dimensions of variation cannot be discerned from other values in other dimensions of the variation, unless that dimension remains constantly set to that value while the other dimensions vary.
4. *Fusion* – The interaction of two dimensions of variation cannot be experienced without experiencing the two dimensions varying simultaneously.

An argument could be made that the approach taken to achieve contrast is a local application of variation, meaning that it views the dimension of the quality of interest, independently of other possible dimensions, as exemplified by Pang (2003). This is a different from the approach of the other three conditions, where a dimension is studied in the context of its interaction with the remaining dimension. Nevertheless, while contrast focuses on a quality being present or not, separation focuses on the entire possible range of values in a dimension. Thus, introducing variation in one dimension by considering different values, serves both as a demonstration of separation for the dimension and contrast for the particular values. Various authors have supported such understanding by providing different perspectives on the same example (Lee, 2008; Lo et al., 2005, p. 21; Marton et al., 2004, p. 16) – one perspective demonstrating contrast on the quality (for example brown), the other one – separation on the dimension (in the example – colour).

In his English language teaching, Lee (2008) investigated an instance of *learning study* from the perspective of professional improvement of teachers. However, in the description of her study, it becomes apparent that the design involved three lessons and only two of the conditions of learning: contrast and separation. Cheng and Ho (2008) designed a learning study on Chinese communicative writing in which they developed variation tasks that aim to involve only separation and fusion. In their book (Lo et al., 2005, pp. 50–56) consider several studies, which selectively involve only some of the four conditions (listed above).

Games and Persuasive Technology

As the review in Section 2.3.2 shows, persuasion is a way to motivate learners to effectively engage with the learning activity. Although, the notion of *persuasive technologies* was introduced by Fogg's *captology* (2009, 2002) relatively recently, persuasive techniques have been widely used for much longer. Games have employed such techniques to improve

player engagement and enjoyment. More specifically, the engaging power of different game-playing media and interactive storytelling approaches are discussed here.

Arguably, one of the factors facilitating learning in games, is the fact that these games support role-play experiences including the associated interactions between real people, despite the fact that these experiences happen in a virtual and fictitious environment. In a series of research publications, Anders Tychsen and his colleagues (2006), investigated the similarities and distinctive properties between different forms of role-play. In a more formal cross-format analysis (Tychsen et al., 2007), compared table-top RPGs, computer-based RPGs and a hybrid format that introduced a game master to computer-based RPGs. From that analysis, the authors concluded that table-top RPGs are the most enjoyable format in general. Arguably, this can be attributed to a combination of discussions of fictitious settings and events between players, as well as face-to-face interaction among other factors. An interesting research question is whether these are determinant factors for enjoyment, and whether their transfer to a different format would improve that medium's enjoyability. It also has the benefit of not imposing a specific paradigm upon user input, thus allowing for creativity and innovation.

The *interactive and digital storytelling* (IDS) community has adopted, as a key driver, the need to develop other means of interaction beyond the ones commonly found in computer games, or in other words, beyond the typical aggressive interactions (like shooting or hitting) or financial interactions (like buying and selling). The widely discussed alternative is dialogue-based interaction. This has been pursued in a number of interactive storytelling systems. A review of early attempts can be found in Ruskov (2005), and some more recent relevant applications are considered in Section 3.1. However, due to the limitations of current *natural language processing* and machine understanding of the related cognitive processes, contemporary interactive storytelling systems are unable to satisfactorily comprehend free-form user utterances (Jurafsky and Martin, 2000). Ongoing research in the area so far has focused mainly on the syntax and ontological meaning of words, as well as pragmatics of utterances through classifying them as dialogue acts. To achieve this, approaches like pattern matching (Mateas and Stern, 2004) and controlled (in other words, restricted) language (Crawford, 2004) are being used. However, it could be argued that with the current level of understanding of human cognition, related to conversations, it is more carefully planned pragmatic trickery that could help build sensible dialogue-based systems, rather than profound ontological understanding of all tacit aspects inherent

in conversations.

Captology defines a functional triad (Fogg, 2002, pp. 23–25) suggesting ways how persuasive technology could be used. The three uses in his classification are as tools, as media and as social actors. It could be argued that this triad could loosely correspond to the three main groups of motivations to play, namely achievement, immersion and social motivations. However, the references in this section show that interaction with peers might be more engaging than its technological alternative. Thus, when designing for a persuasive system, technological *affordances* need to be considered in the context of their non-technological alternatives.

4.1.2 Evaluation

This second part of the methodological review covers evaluation techniques used for assessment of learning, expertise and engagement. Boot, Blakely and Simons (2011) discussed the difficulty of examining the effectiveness of video games for the improvement of cognition. In their summary they suggested that "cumulative evidence suggests a strong relationship between gaming experience and other cognitive abilities, but methodological shortcomings call that conclusion into question".

From this perspective, there are several limitations in this research thesis that I had to work with. The research group at UCL where I conducted this research is domain-specific, rather than having a focused expertise on educational design or technologies. This meant that my work is not part of a wider educational research collaboration; but instead, I had to undertake the educational part independently. This occured in collaboration with my research supervisors who are teaching researchers in the domains of *information security* and *crime sciences*, respectively.

As a result of this, I had to develop assessment materials from scratch and these could undergo only a limited number of iterations, despite the need to continuously improve them. Also, assessment needs to be adjusted to learners. In the context of typical curricula, this is done as part of study programs. In lab studies, participants don't have a shared background and because of assessment design, it is a more challenging task. This is a process that involves both domain-specific and implicit knowledge. Thus, although I attempted to provide a holistic approach, my work naturally considered the task from the perspective of a technology developer, and resultingly I had to rely on my supervisors' expertise in the subject domain.

Learning

Phenomenography studies not the phenomenon itself, but the relationship between the phenomenon and the learner. Phenomenography explores subjective perception of knowledge, or in other words, its ontological assumptions are also epistemological assumptions.

Marton and Booth (1997) made an argument that psychology is mostly focused on what is learned, whereas phenomenography is interested in what and how things are experienced. Thus, it could be viewed as examining the usability aspects of learning experiences. In fact phenomenography examines the *enacted object of learning* (see previous Section 4.1.1), which overlaps with user experiences as studied by usability researchers.

An ideal way to validate learning is through assessment of performance based on that particular learning experience. A point of view to consider when discussing learning assessment is the no-test-is-objective argument, explained in an example by Atherton (2013a). When discussing learning evaluation, or assessment, one fundamental distinction is made between two different forms of assessment: formative and summative. Formative assessment is the kind of evaluation that is used on a continuous basis and is fed back into the learning process. It serves as a form of backward communication about the effectiveness of learning. The other type – summative assessment – is the final evaluation of the learning process that serves as a bottom line of what has been learned in all formal learning.

Summative assessment and final marking are inherent in formal learning and provide an overall measure of learning that can be readily used to assess the employed learning technologies. However, when games are only a small part of the course syllabus the class assessment provides only an indirect measure of the contribution of the technology towards learning. This is why dedicated assessment needs to be conducted, preferably integrated with the overall assessment process of the course.

It should also be mentioned that each particular assessment measures performance only, and not competence. Subsequently, due to distractions or other factors, a learner may score inconsistently over time. One way to address this is by employing a portfolio of assessments. This would naturally lead to a greater demand on qualified assessors, as this has been done in business simulations and reviewed in (Anderson and Lawton, 1992). On the other hand, there is a wide agreement that a portfolio of assessments delivering knowledge profiles of learners – as described by Gijbels and colleagues (2005) – is a good

way to overcome the deficiencies of each singular assessment.

In order to support assessors in the process, different approaches have been taken towards automated assessment. Partially because multiple-choice questions are a popular assessment tool and are easy to automate, they are also extensively used in a number of adaptive learning systems. Multiple-choice questions and their variations have their limitations. Such assessment systems test mostly for the acquisition of *declarative knowledge*, but are less successful in measuring other types of knowledge like *procedural* or *conditional* (see section 2.1). However, the latter two are the main focus of this research effort.

Several authors have argued that different levels of understanding need to be approached for measurement in different ways. The number of these levels vary with different models. For example, Bloom's taxonomy identified four levels, Biggs, in his *structure of observed learning outcomes* (SOLO) distinguished five. Multiple-choice questions have proved an appropriate tool to measure the lower levels of understanding in Bloom's hierarchy (Anderson and Lawton, 1992). One aid in designing measurements of higher levels of understanding is the definition of learning objectives. Examples of how this leads to better focus and accuracy are referred by Gijbels and colleagues (2005).

As indicated in section 2.1, two principal approaches for measuring higher levels of understanding have been identified for the purposes of this research project. One of them is performance in a simulated environment, the other is open-answer questioning. As already discussed in the context of learning assessment, the first technique is very successful when deterministic machinery is involved. However, simulated environments are not the context closest to the real situation when considering interpersonal skills. Role-play has been suggested (Lawrence et al., 1983) as such a similar-to-reality simulation. Furthermore, Lawrence and colleagues showed that people behave very similarly in role play and situations that they perceive as real.

Kickmeier-Rust and Albert (2010) proposed a non-intrusive approach in PBL they called *microadaptivity*. They assessed the learner's mastery by measuring as precisely as possible the progress that had been made towards the master solution, considering every action. They have used microadaptivity for several games designed to teach elementary subjects. However, microadaptivity seems to be too narrowly coupled with the system's ability to anticipate the particular solution that the learner attempts.

The possible applications of Evidence-Centred Design (ECD) (Williamson et al., 2006) is also of potential interest. As part of their research of ECD the Conceptual Assessment Framework (CAF), which consists of five interlinked models, was put forward. Among these five, worth noting are models of the student, the task and the items of evidence that are expected to manifest themselves. CAF has been applied in a serious game for training teachers by Gibson (2007), but it is a relatively novel method, and the degree of its wider success is yet to be seen.

Another group of very popular and successful classic assessment techniques is open answer questioning or essay writing. It has been highly valued because of the freedom of expression (which could be related to autonomy – see Section 2.3 on motivation, – and creativity and innovation) it gives. Open answer questioning also allows for a number of different insights from a single text. This makes it similar to interview questioning, even if not allowing for that kind of drill-down in answers, while still introducing less bias than interviews. Writing is the discussed object for assessment in Biggs' SOLO. However, Powers (2002) demonstrated the current state of the art in automated essay assessment systems can be abused (refer to Section 2.1). These short-comings of automated essay and short answer assessment can be seen as similar to the ones present in dialogue-based interaction (see Section 2.2). Arguably, it shows that because of this complexity, adhering to a formal set of evaluation criteria is too restrictive to be used for text evaluation without a balancing portfolio of assessments. An interesting subsequent question is how text evaluation can be effectively adopted in online systems where text-communication and knowledge sharing are predominant. In order to be able to approach this question, a better understanding of the types of text within each particular environment (and essentially community) is needed. An example of research in user contributions through short text messages can be seen in (Naaman et al., 2010). In a previous study (Ruskov et al., 2010), I reported a similar investigation in the context of knowledge-sharing in games and studied the diversity of written user input.

As already mentioned, ultimately all assessment is about the particular instance of performance: the test. From a situated cognition perspective, as it is described in the substantial subsection above, it is only natural to try to measure actual performance instead of artificial tasks made only for the purpose of assessment. Also, performance assessment is widely used in the business world.

There is some evidence that self-assessment might also be a productive approach.

Falchikov and Boud (1989) made a meta-analytic review of studies and discovered that several factors, including quality of assessment design and maturity of learners, controlled the degree of success of the self-assessment process.

There is a strong opinion in the literature which favours of open learner models (OLM). Research indicated that OLM provides valuable means of reflection that contributes to strengthened and enhanced learner experience (Bull et al., 2008; Bull and Mabbott, 2006; Hartley and Mitrovic, 2002). Hansen and McCalla (2003) provided a possible framework on how learner expertise can be modelled. Research on negotiated OLM, combined open modelling and self-assessment by allowing learners to negotiate their learner model.

Various technologies exist to measure psycho-physiological symptoms. Since intrusive measurement technology could be considered highly distracting, adoption of less intrusive measurements might become a valid assessment component. However, it involves difficulties, similar to the ones mentioned in the context of *log-driven performance assessment* (see Section 2.3) – what is possible to measure is not necessarily something that is informative in terms of understanding the experiences of the learner. This difficulty has been illustrated with eye-tracking technology. It has been used to measure both engagement (Cox et al., 2006; Nakano and Ishii, 2010) and cognitive load (Palinko et al., 2010; Siegle et al., 2008). All of these studies provide some amount of quantitative evidence of correlation (respectively of engagement and cognitive load) with eye behaviour, but do not go on to examine how these reactions relate to learning.

Expertise

Measurement of expertise is very closely related to the measurement of learning as discussed so far. Still, in contrast to learning, where formal examinations are commonly expected in professional communities, this is not necessarily the case. Hence, enquiries with professionals and experts are often made on a voluntarily basis. Alternatively, performance measures used by the employing organisation may be utilised, but these vary with organisations, and it is yet to be seen how any form of sharing of performance data can be acceptable to professionals and experts. As mentioned in Section 2.1 and explained in detail in Section 7.1.2, I contributed to an exploratory study, led by Malheiros, which investigated learner preferences about data, collected in serious games. The results of this study could be used as early guidelines in designing adaptivity and assessment, based on serious games.

Research in professional communities takes an ethnographic approach (similar to

phenomenography) and several iterative and exploratory techniques are used, notably various forms of enquiry, such as data analyses and interviews in Bray and colleagues (2007), Reeves and Malone (2007) and Naaman and colleagues (2010). In the final study for example, the authors analyse more than 3000 messages from Twitter – a *social awareness stream*. They code and classify collected data to observe symptoms, for example that about four of every five *tweets* in Twitter are about the people that write them. To ensure better validity of their data, Naaman and colleagues coded tweets in parallel by two independent coders and reconsidered cases where the two did not match.

Engagement

The design of a serious game needs to be complemented by effective measurements of engagement. Others have investigated a number of other related constructs, including *flow*, *involvement*, *psychological absorption* and *dissociation*, as well as some arguably more distant ones like *presence*, *immersion*, *motivation*, or possibly even *curiosity*. One systematisation of some of these constructs was put forward by Brockmyer and colleagues (2009). The authors argued that the concepts they worked with: *immersion*, *presence*, *flow* and *absorption* – in that order – represent a gradation along a single dimension. For all of these terms, there exist widely used and established self-report surveys that deal with the perceived occurrence of the phenomena.

In a related debate, Slater (2004) – among others – argued that self-report measures are not sufficient and/or accurate enough to measure presence. He suggested that current research on presence has developed a heavy reliance on data gathered from questionnaires. In response, he and others use simulated experiences in immersive environments to conduct their experiments. In a wider discussion, it is appropriate to question whether self-report measures are also adequate for other related constructs. System-collected data is another measure that lends itself, and has already been discussed in Chapter 2, but it could be interpreted only in the specific context and would require dedicated validation.

4.2 Adopted Method

In this final section, drawing from the above review, the methodolody used in this thesis was developed. The overall methodology used here can be broken down into three consecutive phases, outlined in Table 6. The first concept phase is exploratory, applying the evaluation methods to study established serious games. The second design phase is

Study Order	Phase	Type and Purpose of Studies
1	**Concept**	Proxy study to test the evaluation methodology before own game is available
2	**Design**	Formative small-scale lab evaluations to establish a fun game
3		Summative lab evaluation to validate learning and engagement
4	**Production**	Real-world studies with the mature product to establish adoption and facilitation process

Table 6: Development life-cycle for the purposes of design-based research.

concerned with the actual software development of a serious game. Incrementally, in a series of prototypes are implemented to pilot the proposed design and development process and assess the design that has been developed with the mature evaluation methods. The studies in the second part start with a paper-based game prototype and iteratively developed towards a multiplayer online game. The final production phase is concerned with the integration of a matured toolkit into the learning process of university lectures. It involved the development of materials and procedures for the self-paced use of the toolkit as an extracurricular activity.

I initially approached my research questions with an evaluation of existing games in mind. As this was part of an EU-wide research collaboration project called TARGET, the intention was to evaluate the game which was to be developed within the project itself. At the preparatory stages to this development, I started with an exploratory study of a casual browser game called *We the Giants*. This was a case of opportunistic research that allowed me to explore an instance of knowledge sharing with games. It was reported in (Ruskov et al., 2010), and influenced this development with the decision to work with open text user interaction.

The learner audience of this research are young professionals who need to master their responsibilities at work. One such audience that is accessible to university research are graduate students. Typically, they are ambitious (thus somewhat intrinsically motivated) young individuals with a strong interest in a particular field of study. The proxy study in project management was conducted with the students of a Leadership & Communication module. The studies with the *CCO toolkit* were conducted with crime sciences and information security students, both in research and taught programmes.

Within the literature about *threshold concepts*, a recurring related topic is *troublesome knowledge* (Land et al., 2008; Meyer and Land, 2006b). Threshold concepts are the source of troublesome knowledge, but there might also be other pieces of knowledge that are

difficult to comprehend because of the lack of extensive understanding of a related threshold concept. As it was earlier discussed in Section 2. Perkins (2008) explored the link between threshold concepts and proactive learning. The learning domain of this research project started as negotiation for project management, and shifted to the application of an extensive crime prevention framework for the purposes of information security. In the first part of this thesis negotiation was studied, exploring the interplay of stakeholders' *interests* and *positions* in Studies 2, 3 and 4. This is one of the key concepts in that domain and had been extensively studied (Fisher and Ury, 1991), and suggested as a threshold concept (Flanagan et al., 2010). In the development part of this thesis the *Conjunction of Criminal Opportunity* framework (CCO) was embedded in a toolkit (prototype for a *serious game*) for information security training.

The design itself needs to integrate the experience with the serious game into the theory being taught. Following Davies and Mangan's idea (2006), an intermediate step of principles can be derived from the theory before development continues into the design of the actual activities. Learning and engagement need to be aligned so that the attractive game mechanics reinforce the pursuit for understanding the theory, and not divert the interest to, for example self-served collection of points or discussing something unrelated (examples of the latter can be found in Study 1 in Chapter 5). To achieve this I employed principles that encourage learners to explore the object of learning. These principles were based on Yee's (2005; 2012) *motivations to play* and were designed to persuade learners to engage fully with the learning activity. These motivational principles were adopted in parallel with the pedagogical principles of variation as Davies and Mangan (2006) suggested. To achieve this, I employed a process represented in Figure 6, which derives pedagogical and motivational principles from the object of learning, related to threshold concepts within the intended theory. The final design of the serious game and the experience with it, is then guided by the two combined sets of principles.

The research stream which most prominently focused its attention at variation comes from phenomenography, and its proponents have developed a theory about variation in learning (Marton et al., 2004; Marton and Booth, 1997). This is why a common for educational research method in the vein of the phenomenographic *learning study* (Lee, 2008; Pang and Marton, 2007) was used here. This includes the identification of the different perspectives of an *object of learning* and identification of relevant features of that object that are to be subjected to variations. Furthermore, in terms of variation theory, specific

ways to expose learners to *contrast*, *separation*, *generalisation* and *fusion* (Marton and Pang, 2006) are being sought in the design and development of the serious game.

Introducing a social (multi-user) aspect should also be considered as part of this research, because of the benefits of social constructivism explained throughout the literature review. While this is more straightforward with providing a space for learners to discuss and reflect on their experiences, it is more challenging to define a reward system that would achieve a reasonable balance between competition and collaboration in a game. These two together appear to be inherent to both social play and the ludic part of the game design triad.

In order to allow the system to interpret semantics in interactions between users, a formal representation needs to be modelled. Some (possibly trivial) form of simulation can be used to provide users with such an environment. The other two game components of the *game design triangle* – ludology and narrative could also be addressed within game design. Similar to the definition of play, *ludology* could also be related to the introduction of some form of competition (Edwards, 2001) and narrative could be introduced by conveying stories of good and bad practices.

My research had to allow for more clear identification of cause-effect relationships and find a way to get results faster than the typical for learning studies comparison of two subsequent classes of the same course. In effect, an adaptation towards a controlled experiment is used here. The main consequence of this adaptation is that participants are separated into control and experimental groups. The two approaches of experimental lab and class studies are combined.

University ethical guidelines require that learning assessments within studies need to be extracurricular, meaning that they should not contribute to the students' final course marks. In controlled experiments students experience different conditions (in the particular case – one with variation and one without), and it is hypothesised that variation would aid learning, therefore it would be unethical to incorporate marks from the study assessment in the students' final course marks. This obviously holds for *between-subjects designs*, but possibly also for *within-subjects designs*. Nevertheless, as discussed in Section 4.1.1, from the student's perspective, it is preferable that assessment is integrated into a single session, regardless of the aims of its particular parts. Unfortunately, it was not possible to integrate the study assessment with course examinations for this research effort.

Phase	Study (Chapter)	Measured Construct	Type	n	Measures/ Analysis
Concept	**1 (5)**	learning and engagement (negotiation)	class study with a proxy	60*	open questions, MCQ, iGEQ, role-play
Design	**2 (6)**	usability	usability study (formative)	17	usability heuristics
	3 (7)	learning and engagement	lab study (summative)	28	open questions, MCQ, iGEQ
Production	**4 (8)**	engagement	class study	34*	engagement survey
	5 (8)	engagement	class study	25*	engagement survey

Table 7: Studies conducted within this thesis and key information about them. This table was introduced in Chapter 1

* the indicated studies were conducted in a class and not all students provided all types of collected data. Further details on the participating number in each measure can be found in the corresponding chapters.

As discussed in Section 2.1, a portfolio of assessments is beneficial for more accurate evaluation. This is why the adoption of such a portfolio is considered, but it needs to be counter-balanced by a consideration of possible *assessment fatigue*. In effect different measures are administered here, but care must be taken to ensure that these are not too much that they discourage learners from active participation.

In their guidelines for PhD research Herrington and colleagues (2007) suggested focusing on principles instead of variables. Yet, formulating research questions is an important aspect of clarifying the research goals. The three research questions stated in the beginning of this research effort (as declared in Chapter 1) are:

- RQ1: How does experiencing variation in a serious game contribute to improved understanding of learning objectives?
- RQ2: How does experiencing variation in a serious game contribute to improved application of learning objectives?
- RQ3: How does experiencing variation in a serious game contribute to higher levels of engagement with this serious game?

Each of these research questions, requires the identification of assessment measures that would address it. RQ1 needs to measure comprehension, and thus learning assessment metrics that need to be used. As already discussed, depending on the type of knowledge, different measures could be appropriate for each particular type of knowledge. As the main focus of this research effort is on higher levels of learning (conditional, but also procedural

knowledge), it could be measured with contextualised open questions. As a baseline measure, MCQ items are included too.

RQ2 speculates on the application after learning, and thus, an appropriate measure would put participants in a realistic context and ask them to solve realistic problems. As explained in Section 2.1 such are problem-solving as a measure for technical skills and role-play assessment – for interpersonal communication and skills.

For RQ3 different engagement measures and questionnaires could be potentially tried. However, one can argue that the pragmatic and high-level measure of engagement that is most appropriate is the extent to which participants continue to engage in a voluntary activity. Furthermore, introducing one more questionnaire-based measure might introduce assessment fatigue in the context of a learning experiment which that includes several different measures of learning and possibly other surveys.

As a consequence to all these prerequisites, the studies summarised in Table 7 were conducted for this thesis. These follow the process (lifecycle) from Table 6. Yet, the goal of developing a serious game required a range of different studies to be conducted – from the opportunistic Study 1, exploring online data, through to lab usability evaluations, and class studies using games as extracurricular activity. In class studies, I had no control over my sample size, and in the others I had to work with smaller samples due to the many activities that I had to conduct in parallel for the purposes of this research.

Chapter 5: Class Study with vLeader

In my study of an existing game, I went ahead to pilot the evaluation methodology outlined in Section 4.2. To do that an established serious game had to be identified and chosen for the study from the overview of games in project management made in Section 3.1. This game was *vLeader*. It was identified as both relevant to the topic of project management, as focusing on important interpersonal (thus involving troublesome and conditional knowledge) skills and successful according to several studies of learning that showed improvement after the game was used. These studies are summarised in (Sidor, 2008), however none of them addressed deep learning in particular.

In this study the focus is on exploring the effects of variation in response to the research questions. The research questions ask how does the introduction of variation influence learning (RQ1), application of what was learned (RQ2) and engagement (RQ3).

5.1 Method

In the study I conducted to pilot my assessment methodology use of *vLeader* was embedded within a masters-level degree course run at UCL. The class consisted of 60 students from different management degree programmes that were taking part in a class titled *Leadership, Ethics and Communication*. The first 5 weeks of the course were dedicated to the topic of leadership. Use of *vLeader* was embedded in that period. In addition to the game, the course had a number of other activities dedicated to the topic of leadership including lectures (reviewing different theoretical models of leadership), case studies (shedding light on practical complexities) and videos (demonstrating leadership behaviours in action).

As a result of the methodological review conducted in Chapter 4 for this study I used a modified methodology for the evaluation of learning effects in serious games. I chose to adapt the *learning study* (Pang and Marton, 2003) in order to evaluate a serious game. As a research methodology the *learning study* is an attempt to take advantage of, and combine, controlled experimental studies with an incremental improvement method called *lesson study*. The learning study is used to evaluate by comparison of learning effects between subsequent deployments of a course, introducing improvements over several iterations. In order to fit within a single university term I attempted to provide two comparable learning conditions within a single class.

The study used a *between-subjects design*. The independent variable was variation, which included the conditions variation (the *experimental group*) and no variation (the *control group*). At the beginning of the study, the students were randomly assigned to either the experimental condition or to a control group. I did not have the necessary level of control of the software to implement the separation between control and experimental conditions, so it was implemented through the paper-based materials (containing different written instructions for the two groups) for each scenario, called *activity sheets*. These are described in detail in Section 5.2.

5.1.1 Learning Environment

To put it in context, *vLeader* was analysed according to the evaluation framework of de Freitas and Oliver (2006) as was reflected in Table 3. Here it is considered in further detail how it is meant to be used and how this was adapted for the purposes of this study. This is done by reviewing the four dimensions of the framework: context, learner specification, pedagogic considerations and mode of representation.

The simulation game was developed by Simulearn Inc. and is commonly delivered as the main learning resource for commercial webinars, which defines the context according to de Freitas and Oliver. These webinars are typically a series of online conferences before and after learners play each scenario at home (a take on briefing-debriefing of the game experience). At a webinar a facilitator introduces features of *vLeader* and encourages learners to discuss their experiences with it. Learners are encouraged to play each scenario eight times. The game itself is described in Section 5.2.1. The webinars are accompanied by a Student's Workbook which guides learners through the learning process for each of the sessions. Simulearn Inc have also occasionally deployed the software in university classes. In doing this, they have identified the need to break down their Student's Workbook into smaller bits of information that would allow for better adaptation to the curriculum of the host organisation. However, this process is still at its start and was piloted with the materials prepared for this study (see Activity Sheets described in Section 5.2.2).

Using *vLeader* in commercial trainings, as opposed to university classes, implies that there's very little common ground between learners across courses. Thus it is difficult to talk of general learner specification, as de Freitas and Oliver suggest. Yet, within a single webinar typically learners come from the same company which gives them certain common background. The actual simulation game being single-user, discussion and reflection are the activities that allow for group work and this is being utilised in the

webinars and university classes, including in this study.

With respect to pedagogic considerations, *vLeader* is a typical example of a simulation game promoting experiential and exploratory learning. It employs techniques like conflict resolution (a special type of problem-solving) and role-play with changing perspectives. Yet, as Kebritchi and Hirumi (2008) point out that it is typical for educational games, the authors of the game do not elaborate whether and how they derive the possible game activities from a pedagogic theory (Aldrich, 2003). Despite that, the authors manage to come up with their own theoretic model and a game design that allows parallels to other leadership and negotiation theories like Fischer and Ury's model (1991), even if this has not yet been documented in corresponding learning materials. For the appropriation for the purposes of this study the game (detailed in Section 5.2.1) and class (see learning objectives in Section 5.2.2) goals were elaborated and related to one another.

In *vLeader* the authors have chosen to use a first-person 3D environment, supposedly to help players identify with the played role. They have also very cleverly adopted an interaction paradigm that is closed in itself (meaning that there's only a limited number of possible interactions), but still allows for great expressive power that is relevant to leadership and negotiations. This is further elaborated in Section 5.2.1.

5.1.2 Procedure

The study consisted of three major stages: learning, assessment and post-study interviews. The learning stage was influenced by the available game scenarios, the experimental between-subjects design and the procedure for handling each of the scenarios. The learning stage occurred over 5 weeks. During this period, five classroom sessions, 180 minutes each, were used for teaching. In each of these sessions, only a part was dedicated to the study in order to introduce the game scenarios, to administer study materials, and to facilitate group discussions around the game playing experiences. The rest of the classroom sessions were used for delivering lecture materials on the topic of leadership. The game playing experiences took place individually in between these classroom sessions. As was the practice in Simulearn Inc's courses, it was suggested to students to play each scenario at least 8 times. Before each class the corresponding learning materials were made available on the course learning support system (Moodle).

In Week 1, the students were introduced to the game and the study. Students were given 10 minutes to do the pre-test questionnaire. The course tutor then continued with the course

lecture. Towards the end of the 180 minutes, the first game scenario was introduced and the students were each handed a paper-based version of the activity sheets. Students were then encouraged to play Scenario 1 of the game at a time of their own choosing prior to the next classroom session. This introduction to the game during the first week was disrupted by a fire alarm and subsequent evacuation of the whole building. As a consequence the introductory session was less elaborate than originally intended, but the study had to continue, effectively relying on the online presentation and follow-up student inquiries.

In Week 2, the 180-minute classroom session incorporated a 10 to 15 minutes group discussion that focused on the students' experiences of Scenario 1 of the game and relevant leadership theory. Typical questions in these discussions were addressed to how each of the game characters behaved and whether students could relate these game experiences to real life examples. Towards the end of the session, I introduced students to Activity Sheet 2 (provided online) and Scenario 2. Again, the students were encouraged to play the game prior to the next 180-minute classroom session.

During the remaining weeks, a similar pattern was followed: the game scenario introduced the previous week was discussed in a short 10 to 15 minute session and, at the end of the class, the next game scenario and activity sheet were introduced. The initial discussion was held in relation to the relevant leadership theory presented in that particular class. Due to the restricted number of classroom sessions dedicated to leadership, it was necessary to introduce and encourage students to play scenarios 2 and 3 in one week. These two were chosen in particular because of the similarity between the two situational contexts. In the final session (Week 5), following a short discussion of the fifth and final game scenario, the post-trial written assessment test was administered.

5.1.3 Learning Assessment

As explained in detail in Chapter 4 written assessment tests are a widely used form of assessment in studies of learning technology (Anderson and Lawton, 1992). Free-form written assessment methods can be used to measure deep learning (Biggs and Collis, 1982). This study employed written assessment tests before and after exposure to *vLeader* as a measure of learning that took place.

For this study, a bespoke written assessment questionnaire was developed to measure understanding (RQ1). It consisted of 3 open-ended questions and 7 Likert-scale questions (see Section 5.2.3). Responses to open-ended questions were examined via content

analysis. In order to test application (RQ2), a bespoke role-play assessment was developed, putting students in a situation, similar to the ones they encounter in *vLeader*. Retention – the number of times that learners return to play the game – was intended to be used as a measure of engagement (RQ3).

After the learning and assessment parts of the study were completed, a series of in-depth semi-structured interviews were conducted. The aim of these interviews was to further explore students' perceptions of the game and application of variation in particular. The planned questions are listed in Table 8. These were adapted to become relevant for the experience of each individual participants. For example, changes considered where students used *vLeader* and how this affected their experience, when a student did not engage with the game or the role-play, they were asked what made them not engage, instead of asking for comments on the experience.

5.2 Materials

Four sets of materials were prepared. The game *vLeader* itself was one of them. Activity sheets, based from the game-accompanying learning materials were a second set of materials. The learning assessment tests were another set of materials, and the final one was the materials prepared for the role-play assessment. All these are described below.

1	How did you find vLeader? Did you find it to be a good use of your time?
2	Any particular difficulties with it? What about the game installation or the game interface? Do you consider that playing on the library computer cluster rooms influenced the way you played it?
3	What did you think about the game visuals/graphics?
4	What do you think of the freedom to choose when to exercise with vLeader (as opposed to fixed times, e.g. as you did with MarkStrat in a different module)? Do you want to comment on the duration of a single scenario (around 15 minutes)?
5	What do you think of the red-green interaction interface (the positive-negative sliders for characters and ideas)? Did you find it restrictive? Were there any specific other interactions that you wished were present? Did the red-green interface realistically model real situations?
6	How would you describe the differences between practice or advance mode? Did you have any criterion on when to switch from practice to advanced mode?
7	Did you discuss your game experience with someone? Would you have discussed it online? Would it have been better if you were allowed to comment anonymously?
8	How did you use the activity sheets? Did you use the sections selectively, if so, how? Were they presented in an efficient way?
9	In one sentence, please describe what you learned from the game. Would you take part in another study like that?

Table 8: Questions asked during the in-depth interviews at the end of Study 1. These questions are indicative and were adapted for each individual student.

5.2.1 vLeader

vLeader is a simulation game that aims to provide a practice environment in the domain of leadership skills (Standifer et al., 2010). The *vLeader* game embodies its own theoretical framework which also informs the design of the game. An introduction to the framework is available to players via the instructional materials that the learner can access through the menus of the software. Although this framework was not considered important for the purposes of the course, there was no way to restrict students' access to it during the study. Therefore students were told that they are welcome to explore it themselves if they wanted to, but they should consider it only as one of many possible theoretic frameworks about leadership.

The game provides learners with role playing experiences within a series of simulated business meetings (see Figure 7). The game developers argue that business meetings are prototypical situations for practising leadership. The game dynamics are based on three variables: the player's influence, group opinion towards the player and tension in the meeting. Personal influence determines the power of player's position. When the player suggests a new idea, they put their personal authority at stake. If the idea gets approved,

Figure 7: A screenshot of Scenario 4 of vLeader, showing a scenario setting, subtitles (blue area above), red-green opinion sliders and blue idea progress indicators. Ideas listed to the left are ones that are not currently brought to discussion. Ideas on the right are the ones that have already been passed (agreed).

their personal authority increases. Group opinion represents the attitude of participants in the meeting towards the player. If a player manages to find an effective balance in their interactions with the characters in the meeting, the opinion towards the player improves. Finally, tension is measured through player's ability to manage the conflicts in the meeting. Players get the opportunity to review their performance on each variable at the end of each scenario.

In *vLeader*, players interact with the game by exchanging positive or negative signals towards character or ideas. Ideas in *vLeader* represent topics that are being discussed. Progress in the discussion for a particular topic is represented by a progress indicator (see the blue progress bars in Figure 7). Players can send a positive or negative signal for a particular idea or a particular character using clickable opinion sliders (also seen in the bottom of Figure 7). Each slider is coloured in a red-green gradient. If the player left-clicks on the green side of the slider for a particular character, their avatar makes a positive comment. The comment is negative if they click on the red side.

Thus interactions in *vLeader* are defined according to three dimensions:

- valence (positive-neutral-negative) – because of the continuous values that the red-green sliders can take, the strength of this dimension formally could be uniquely mapped to the number interval [-1,1].
- target (towards people or towards ideas) – a binary relation which corresponds to a target object.
- Verbal/utterance or non-verbal/gesture – in *vLeader* non-verbal communication is manipulated like verbal, but without taking turn in the conversation

A fourth dimension is the actual textual content, if present. However, *vLeader* does not consider the actual text in its game mechanics. Thus the three dimensions define each interaction act (single piece of interaction) by the player.

The exact utterances that emerge throughout play are limited and can often be unrealistically repeated. Simulearn Inc. argue, in accordance to ideas of variation and invariance, that this issue is of secondary importance and helps students not get distracted from the main learning focus, being the intentions behind each particular utterance. They have taken a decision to restrict the interaction interface to the intention of what is being said. Simulearn Inc. argue that in leadership situations, it is not as important what exactly

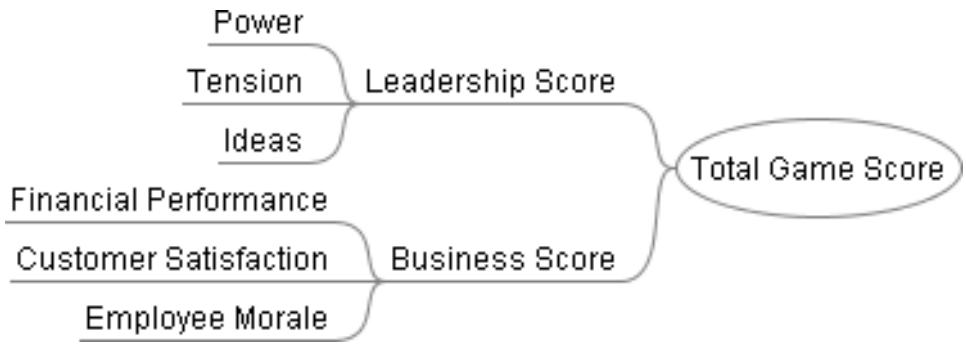


Figure 8: Components of in-game performance scoring and their aggregates. When scores in vLeader are aggregated, an average of the corresponding components is taken.

is being done, but rather why is it being done. The right mouse click has a similar meaning, but instead of comments it represents non-verbal behaviours.

The game scoring mechanism consists of 6 values (see Figure 8). These values are grouped in two components: 1) *leadership score*, which combines *power*, *tension* and *ideas* sub-score and 2) *business score* based only on passed ideas in the scenario, which consists of three sub-scores *financial performance*, *customer satisfaction* and *employee morale*. These are the only form of immediate feedback that students get for their performance and thus they are expected to influence player behaviour in the game.

vLeader consists of five game scenarios. Students get access to play a scenario only when they pass all previous ones. The first scenario is designed to get the player used to the interface and involves managing one subordinate. In the second scenario, the player is required to manage two subordinates who have a latent conflict between them. The third scenario puts the player in a meeting with both superiors and subordinates in the organisation. The fourth scenario represents a situation where the player has the least formal power in the meeting. The fifth and final scenario features a crisis situation of the company management, after the occurrence of a risk event, and the player is only one of several attending managers. Regardless whether in possession of formal authority or not, the player's agenda is to get certain ideas agreed by the group.

A last aspect of the design of *vLeader* that is worth mentioning is that the simulation game provided students with the opportunity to choose the mode of play that they want to try: *practice* and *advance*. As the name suggests, practice mode was intended to be used when exploring a scenario. It featured a continuous graph displaying the development of the conversation variables described above as a form of immediate and continuous feedback. Advance mode, on the other hand does not show this information during play. Users have to play a scenario at least once in *advance* mode, in order to progress towards the next one.

Arguably these design decisions make the two modes correspond to Malheiros and colleagues' (see Section 7.1.2) recommendation for separate learning (in this case practice with immediate feedback) and assessment (*advance*) modes.

5.2.2 Activity Sheets

Typically *vLeader* is delivered together with a Student Workbook. In this workbook, there is a strong coupling between the simulation and the corresponding theory. Within the current study, the workbook was considered too restricting by the course tutor. So, for the purposes of the study, I developed a set of dedicated *activity sheets* – these adapted information from the Student Workbook to make it more relevant to the course requirements. The activity sheets were approved by Simulearn Inc. and the course lecturer. The activity sheets also provide a means of introducing variation into the game experience. Despite the fact that they were called activity sheets, as the name was suggested by Simulearn Inc, these documents were 4 to 5 pages each. Including copies of background information from the game scenarios figures and blank spaces for student reflection (see in Appendix B).

The activity sheets consisted of six sections: learning objectives, scenario background, business scoring tables, goals, hints and reflective questions. The sections were clearly distinguishable (compare to *information chunking* as reviewed in Section 2.3), providing students with the opportunity to use them selectively, according to their own preferences.

The *learning objectives* were one-sentence descriptions of the intended learning outcomes for each particular scenario and represented the connection of each game scenario with the particular lecture. The *scenario background* section provided students with a written description of characters and ideas in the scenario. These were directly transcribed from the background information texts, delivered by the game in order to make the information more accessible. The *business scoring tables* provided a transparent scoring mechanism for the value of passing each particular idea. They represented a balanced scorecard of business score points that players receive when they pass certain ideas within the game. They were thus crucial towards high performance along the business score component of the game performance scoring.

The experimental and control conditions were implemented via the *goals* and *hints* sections of the sheets. For the control group, the goals and hints were directly transcribed from the *vLeader Student Workbook*. The goals and hints for the experimental group were

amended for Scenarios 2 to 5 so that they focused more narrowly on one specific aspect of leadership each. Marton and Pang's (Marton and Pang, 2006) conditions of learning were put into practice in *vLeader* as follows:

For Scenario 1, a decision was taken not to introduce variation because the original scenario already employs *contrast* in encouraging students to perform different styles of leadership in order to be able to compare them. Students were asked to be directive, participative and delegative in playing subsequent games. This was intended to allow them to compare how different styles influence other participants in the meeting.

In Scenario 2, variation was introduced through the hints section by focusing on the original concept from the Student Workbook and allowing for *separation* of time planning. The original text focused on keeping in mind the end goal and planning for preparatory work that could lay the path towards it. The introduction of variation to the other group, on the other hand, asked students to initially try to directly aim for the final goal, and in subsequent play to try to plan for sub-goals by way of preparatory work. When instructed not to plan, the intention was that even those that were naturally inclined to do it, would deliberately postpone such an activity until after having played. This approach was intended to underline the difference of whether to plan before a meeting or not.

Scenario 3 explored different approaches to the conversation with regard to who dominates it. The original activity sheet suggested first dominating and then letting others dominate, as this was done in all previous scenarios. The introduction of variation provided goals that suggested supporting someone else in the conversation and then subsequently striving for a better balance. This allowed students to clearly distinguish the effects of taking sides in a conversation. The focus of variation in this case was on *generalisation* on the introduction of personal bias in a conversation.

This found its continuation in Scenario 4, where it was intended to allow for clearer *generalisation* on the role of personal influence, specific suggestions for who to liaise with. Whereas the original instructions suggested "building an alliance with one or more characters for a strategic purpose", generalisation was strengthened by explicitly suggesting to ally consecutively with the two different characters that opposed each-other to a strongest degree. This allowed for more controlled and exhaustive variation, focusing on the two opposing sides in the conversation.

Finally, variation in Scenario 5 was intended to allow for clearer *separation* of

performance from business results. To do that, students were encouraged to aim for as balanced business score (regarding its components) as possible. The intention was that while they still play for high results, they should aim for balanced business score components, which would allow the power and tension scores to be separated from the idea-related business scores.

The last section of the activity sheets was designed to encourage students to reflect on what they had learned in the scenario. This included questions that students were asked to answer before playing in order to plan for their success and others that were intended for after playing as a means of retrospective reflection. The success planning involved an engagement strategy for each meeting. It focused on intended ideas to be passed on and balance between signals sent to people and ideas. The reflective questions concerned satisfaction with results, what styles of leadership were used and what parallels to reality students could make.

5.2.3 Pre- and Post-Study Written Tests

These tests were developed in collaboration between the researchers and the module lecturer. A mix of open-ended questions and fixed-response questions (MCQ) were used to try to capture different levels of learning as planned in Chapter 4. The students were given 10 minutes for the test, so open answers had to be short.

The first three questions (Q1, Q2 and Q3) were open-ended. Q1 aimed at capturing the respondent's general conception of leadership and asked for an example. This question was intentionally formulated in terms of personal understanding rather than definitions. Nevertheless, it measured *declarative knowledge*, thus clearly addressing RQ1.

Q2 presented a situation that students were expected to be familiar with. It depicted a situation in which the student was part of a team that had to deliver, but there is tension within the team, a theme covered by Scenario 2. The experience of variation in Scenario 2 was expected to lead to a greater awareness of the need to break down solutions into sub-goals. Those who experienced variation in Scenarios 3 and 4 were expected to recognise their own role as only one contributory factor in collective decision making, rather than as an individual decision maker. Finally, students who experienced variation in Scenario 5 were expected to consider all three aspects that corresponded to the business score: financial performance, customer satisfaction and employee morale. Similarly to Q3 below, this question asks about application (RQ2), but for only a written answer, which limits the

depth and validity of the possible response. For this reason, application was intended to be actually measured via role-play, described in the subsequent section.

Q3 was intended as a crisis situation, as it appears in Scenario 5 of *vLeader*. The corresponding pattern could be described as a risk event having occurred and giving rise to two consequences – one more apparent and one more urgent. Students were required to explain how they would resolve it. Similar to the previous question, the experience of variation in Scenarios 3 and 4 was expected to lead to greater appreciation for collective decision-making and variation in Scenario 5 – again appreciation for the corresponding three business aspects. Q2 and Q3 measured *conditional knowledge*.

Q4 included 7 statements about leadership: students were required to indicate their level of agreement om a Likert style response scale with five options from 'agree' to 'disagree'. For four of the statements (Q4i, Q4iii, Q4iv, Q4v, Q4vi and Q4vii) it was expected that students in the experimental group were more likely to agree. For one statement (Q4ii), it was expected that students in the experimental group were more likely to disagree. All of these were considered to measure *declarative knowledge*.

As part of the assessment design, I prepared sample correct solutions for the open-answer questions, but the diversity of answers provided by students in the study (see results) meant that I could not use the sample solutions as a baseline for assessment. This is a commonly observed effect, as reported by Bransford and Schwartz (1999, p. 7) for example.

5.2.4 Role Play Assessment

The role-play assessment involved a 15-minute pervasive experience with two actors and a number of cross-media artefacts that were used to provide the role-play background. The experience was based on a 2-page outline of an interactive script that was iteratively developed over a series of rehearsals and pilot role-plays. Professional actors were recruited for an improvisational performance with the constraint that separate performances should be as consistent as possible in order to allow for subsequent comparison of students' behaviours.

The plot involved a crisis situation within a small training company where a keynote speaker for an upcoming event has made a late cancellation. Students were invited to a meeting to resolve the problem. The resolution itself did not present a great challenge, as possible alternatives were suggested by the actors upon request. The leadership challenge consisted in being able to overcome the in-role tension between the actors and to get their

buy-in on the resolution plan and what each of them had to do to realise it. The plot was that when students entered the room, the actors were already in a conversation about the problem.

Students had the story background delivered to them through a fictitious promotional website of the event. They were led to it through a combination of a mail informing them about the occurrence of the risk event and a simulated old job offer flyer, not delivering any new information and announcing their in-role position. During rehearsals, actors developed strategies on how to fill-in gaps that students might have in their background information, effectively waiving any requirements on students' preparation, before attending the meeting. The actual role-plays took place in an informal meeting room with a coffee table and soft chairs. A clock and two cameras facing the student were installed.

The role-plays were recorded and analysed with the purpose of making them comparable to the experiences that students had with *vLeader*. To achieve this, the video recordings were coded for students' contributions to the conversations (utterances), behaviours and events, which commonly do not depend on students, but are used to set the scope of each particular role-play. The first two types of annotations were coded according to the *vLeader* interaction framework (positive-neutral-negative, people-or-ideas, verbal-or-non-verbal). In order to make the coding process more manageable and to avoid unnecessarly complicated analysis, I decided that only the discrete subset {-1, 0, 1} of the positive-neutral-negative dimension (-1; 1) would be considered. Not all interaction acts match the *vLeader* interaction framework. The ones that did not were classified in two other categories: they were considered either inevitable or inefficient. Inevitable were acts which signify utterances or behaviours that were driven by the rest of the conversation (like response to a question by one of the actors or a gesture accompanying something that is being said). Inefficient are acts that do not match the *vLeader* interaction paradigm. This includes distracting behaviours, unfinished sentences, etc. but also asking for information that students should already be aware of, acts towards people, but not directed towards one person in particular and so on. The thus-coded interaction acts were aggregated and compared to students' results in the game.

5.3 Results

The results of this study are reported in five sub-sections: student involvement, in-game performance scoring, answers to the written assessments, preliminary analysis of role-play assessment and subjective student experiences, shared in post-study depth interviews.

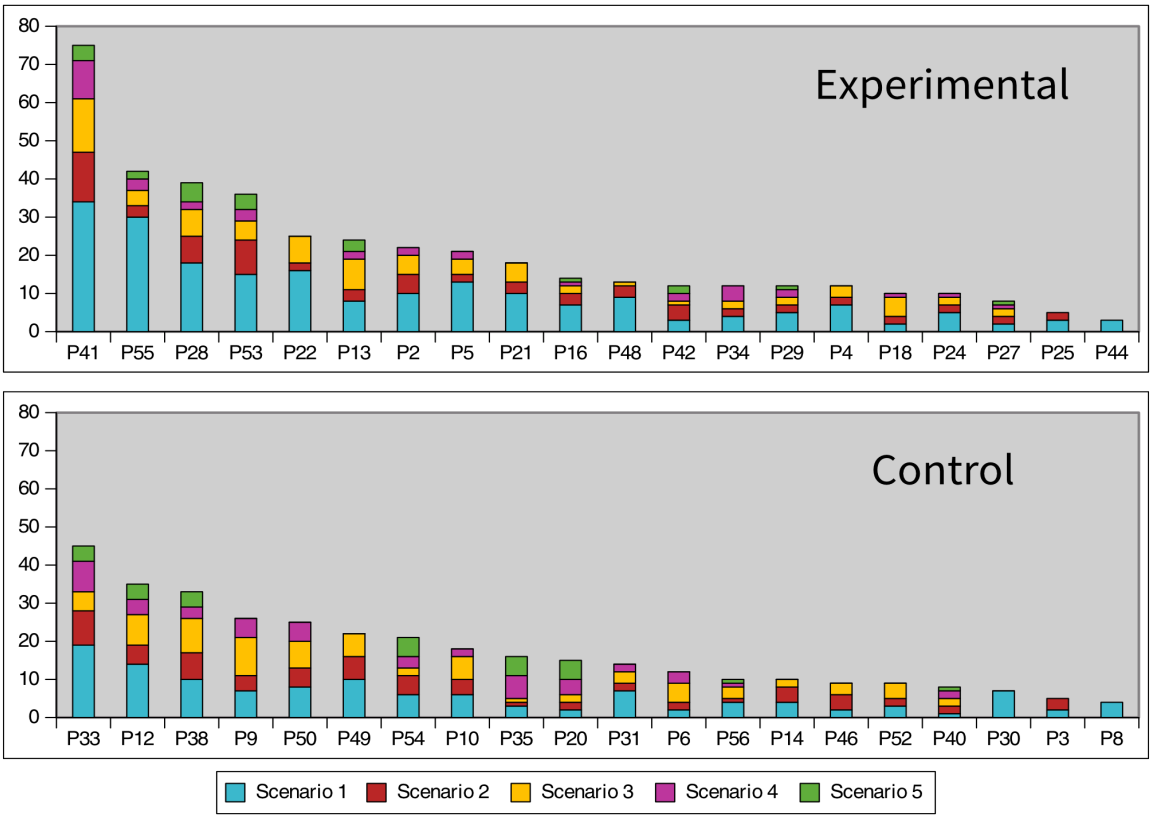


Figure 9: Number of games each participant played – colours indicating scenario number. Student pseudonyms, as used to identify other sources of data are provided. The difference between the two groups is not statistically significant.

5.3.1 Engagement

Students were given access to university computer pools to play the game there if they wanted. Out of all 60 students, 17 reported doing that. A common reason was that their private computers were incompatible with Windows, which was the only platform that *vLeader* supports.

While some students played the game quite actively, a number of others did not engage with the game at all, thus there was no data collected for 14 out of the 60 students (23%).

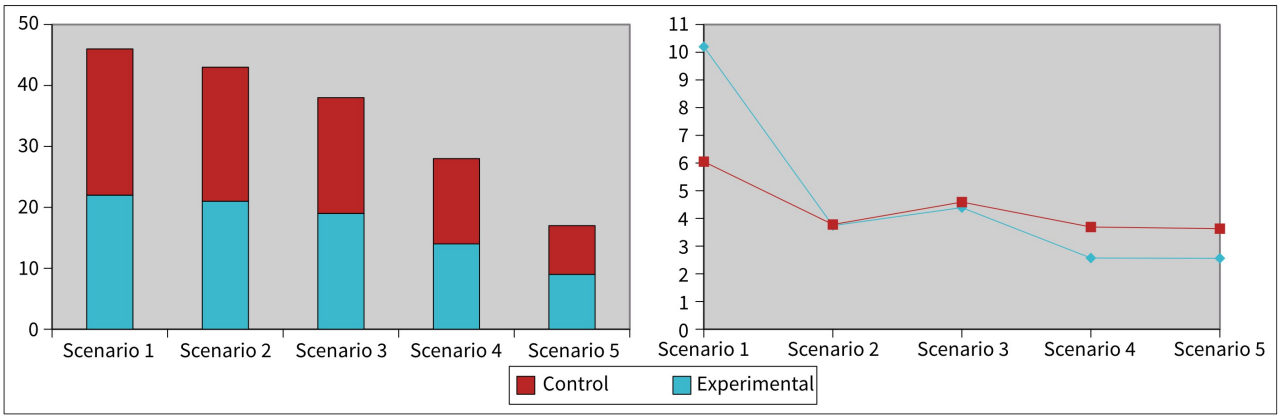


Figure 10: (a) Number of participants that played each scenario in the bar chart to the left; and (b) Average number of plays per participant, according to scenario and experimental group in the line chart to the right

Throughout the period of engagement, students showed gradual reduction in their involvement with the game, both in terms of scenarios played (see Figure 10a) and number of plays per scenario (see Figure 10b). The number of students that played the game fell from 38 for Scenario 3 to 28 for Scenario 4 and only 17 for Scenario 5. Over the prolonged use of the simulation game, students seemed to prefer to play between 3 and 4 games per scenario, despite the recommendation to play each scenario about 8 times.

Both from Figure 9 and Figure 10b it is visible that the experimental group engaged with Scenario 1 more than the control group. Results from statistical testing in Table 9 show that this difference was statistically significant with threshold ($\alpha = 0.05$). However, this cannot be interpreted as an indication in support of greater engagement in the case of variation (RQ3), because at this point of time variation was not introduced yet. There was no such difference in the later scenarios.

Since some students (23%) did not play the game at all, it was decided that a third group should be formed for the purposes of the analysis. This non-player group is treated as an additional group in the analyses below. In order for variation to work, students had to have as a minimum requirement, several different experiences with the game. For this reason, it was decided to include in the non-player group also those students who played the game just a very few times. This was defined as students who played Scenario 1 up to two times and subsequent scenarios no more than once. As this group was not created through random allocation, but self-selection, the results related to this group should be treated with caution, for example because. there is the possibility that differences are due to factors other than game play.

Three different measures of student performance were used in the study. These were performance data from the simulation game, pre- and post-study written tests and post-study role-playing scenarios.

5.3.2 Performance Data

The investigation of the validity of *vLeader* performance data allowed for one direct measurement of effects of variation and lead to three different observations, related to students' experiences within the study.

In contrast to the rest of the scenarios, the way variation was introduced to Scenario 5 and the data that *vLeader* collected, made it possible to directly measure the effects of introducing variation on student behaviour through the game scoring mechanism. In order

Plays	Experimental		Control		t-test	
	Mean	Variance	Mean	Variance	t(df=38)	p
Scenario 1	10.2	78.59	6.1	20.58	1.86	0.0364
Scenario 2	3.6	8.89	3.4	5.52	0.18	0.4304
Scenario 3	4.0	11.00	3.9	9.04	0.05	0.4802
Scenario 4	1.8	5.12	2.4	5.62	-0.82	0.2090
Scenario 5	1.2	2.66	1.5	4.37	-0.51	0.3079
Total	20.7	282.45	17.2	121.54	0.77	0.2241

Table 9: Comparison between the control and experimental groups in terms of number of plays for each scenario and total for all scenarios. Unpaired one-tail t-test was used. Number of plays differed only in Scenario 1, as indicated in Figure 10b

to do that, values of variance between different components of the business score was compared in both groups, applying statistical significance tests on the *variances* between the score components. The experimental/control group was used as independent variable and the variance – as a dependent variable. Using unpaired t-test with the assumption of equal variances no statistically significant difference was found. The dataset for this test is available in Appendix B and results from a one-tail test were not significant ($t(df=50) = -0.0322$, $p = 0.48$). This result indicates that the activity sheet for Scenario 5, intended to introduce variation for the purpose of improving understanding, did not lead students to exhibit wide variance in their performance. Such variance would have been a sign of variation in the score, when asked to explore variation in the assignment.

Three observations could be made considering students' perceptions of *practice* and *advance* modes, as they were provided in *vLeader*, interference of exploratory behaviour in performance data and different tactics students employed in order to conceal their scores. The first of these observations involved students' perception of different modes of play (*practice* and *advance*). In line with Malheiros and colleagues' (2011) findings about user expectations, some students wrongly assumed that while playing in *practice* mode, their performance data was not being recorded for later review. This was not suggested in the consent form, and I communicated the misunderstanding to all students once I became aware of it.

The recorded performance data was explored to verify this phenomenon and results are aggregated in Table 10. To ensure validity, only those data entries were considered, that belonged to students who had played a given scenario in both *practice* and *advance* modes. These entries represented 32% of all collected game data and 93% of students that

Score Component	Aggregated in	t(df=121)	p
Power	Leadership	-2.78	0.0031
Tension	Leadership	-0.60	0.2764
Ideas	Leadership	-2.21	0.0146
Financial	Business	-3.66	0.0002
Customer	Business	-3.76	0.0001
Employee	Business	-2.34	0.0106
Leadership	Total	-3.35	0.0005
Business	Total	-3.70	0.0002
Total		-3.86	0.0001

Table 10: Probability thresholds for differences in vLeader score components, according to mode of play – practice or advance. The figure was calculated with a one-tail paired t-test. Compare the score hierarchy with Figure 8.

played at all. Although it was considered that the actual student and scenario number might have a mediating/moderating role, this was ignored and paired t-test was conducted with mode of play as independent variable and score components as dependent variables. A statistically significant difference emerged over several sub-component scores and over all aggregated scores. There was a higher confidence as far as business score components are concerned. Students had clear insight of how business scores were being calculated. As a consequence, it could be speculated that they were better able to affect these, when trying to reach a high score in *advance* mode. The statistical analysis shows that students' intention to perform better when playing in *advance* mode had shown least results on the tension score. They were least able to purposefully demonstrate their ability to achieve a good balance of tension when they attempted to do this.

The second observation had to do with the fact that in order to encourage exploration and variation, during class discussions, students were asked also to experiment with being passive in the game, or taking different sides in an argument (for example see Scenarios 3 and 4). As a consequence of this, and sometimes driven by their own intention to explore alternatives, in certain plays students would have reached scores that do not reflect the maximum of their potential. Although in the case of most students, it could be argued that they did this exploration in *practice* mode, the fact that several students played in *advance* mode only raises questions about the validity of these results as a measure of overall proficiency. As a reminder, notice that not playing in *advance* mode does not allow for progression to Scenario 2 and beyond.

The third and last observation that had the effect of undermining the validity of collected performance data was made, based on reports by some students that they had figured out ways to selectively submit their results. One student in particular reported that shortly before finishing playing a scenario, he would determine whether he was satisfied with his performance, and if not, he would quit the game before the final scores are calculated and submitted. In this way he had developed a new leverage to determine which results he wanted submitted, and which – not. Two other students reported playing the game disconnected from the Internet, and thus inhibiting it to submit results. While such uses of *vLeader* were envisioned when Simulearn Inc. developed the game, the contingency mechanism for later submission works only when users willingly collaborate and submit the locally stored data manually. Yet, the collected data suggests that these were only exceptions that should have only minimal impact on the overall performance of the groups.

5.3.3 Written Assessment

Pre- and post-study written tests were used to measure student learning. Of all 60 students on the course, 57 filled in the pre-test assessment sheets. 50 filled in the post-test assessment sheets. A total of 47 students managed to complete both the pre and post-test written assessment. This resulted in the following sample sizes for each of the conditions: 17 students in the experimental group, 16 students in the control group and 14 students in the non-player group. This section presents the results of the analysis of the responses to the written assessment questions. The complete responses are provided in Appendix B.

The first three open-ended questions were analysed using content analysis. Written answers were examined in order to design a bespoke coding scheme that would capture all content within the answers. After that each response was coded with a set of binary codes for each question. The unit of analysis for Q1 was words and for Q2 and Q3 it was sentences. Non-hierarchical coding was used, meaning that certain content could be

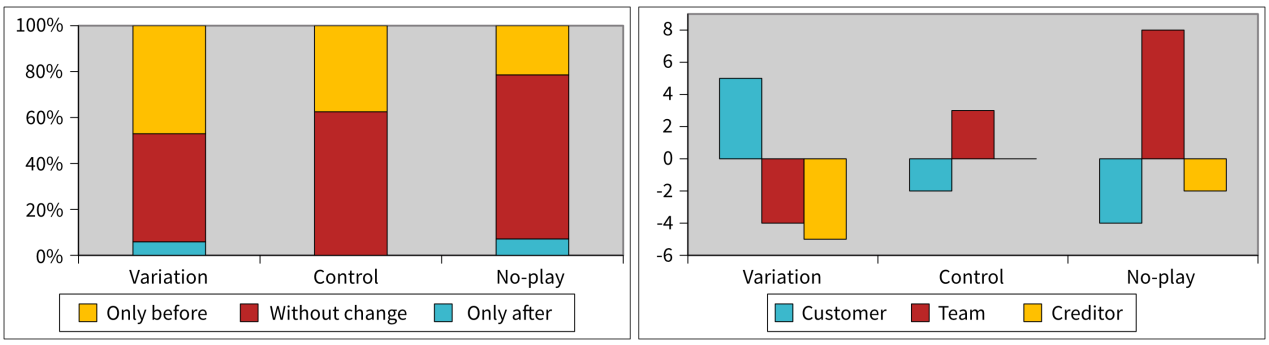


Figure 11: Change in students responses from pre- to post-test sessions in the bar chart to the left(a) employing influence to explain leadership in Q1; (b) mentioning particular stakeholders in response to the situation in Q3 in the bar chart to the right.

assigned several codes.

There was a noticeable change from the pre- to post-test session in the students' answers to the open questions with students providing shorter answers in the post-test sessions. This was discussed in the depth interviews. Further findings are outlined in Table 11.

Words were used as unit of analysis of Q1. Analysis of the vocabulary, used by students' responses to Q1, showed that the percentage of student responses talking about *influence* increased from 15% at the beginning of the study to 47% at its end. This change was stronger in the variation and control group (with 47% of students in the variation group using *influence* to describe leadership in the post-test, but not in the pre-test) and weaker in the non-player group (where there was such change in 21% of the responses). Percentage representation of changes can be seen in Figure 11a. This is a weak indication of support for RQ1.

There was also a noticeable drop in number of students that provided an example in their answers as it has been requested. Whereas at the beginning of the study 62% from the respondents provided some form of an example, at the end of the study only 38% did. One possible explanation for this could relate to the decrease in student engagement by the end of the study.

Q2 and Q3 were problem cases (addressing RQ2) and thus the chosen unit of analysis was

Question (unit of analysis)	Finding
Q1. How do you understand leadership? What example could you name? (word)	Mentions of "influence" in responses showed a 32% increase when describing leadership from 15% at the beginning of the study to 47% at its end. A 24% drop in examples of leadership provided – from 62% to 38%. However, there were no statistically significant differences across the experimental groups.
Q2. You are working together with two other colleagues on a project. One of these colleagues regularly skips meetings and seems distracted when you discuss the project. His work has been of poor quality. What would you do? (sentence)	No observable tendency or other findings.
Q3. You work in a start-up with two other people. A major client has delayed a planned payment and as a result the company in turn cannot make a payment due by the end of the day. Your colleagues start an argument whether the company should stop working with that client. What would you do? (sentence)	The group subjected to variation shifted its focus of attention towards the customer, whereas the control and non-playing group shifted their focus in the other direction: towards the company.

Table 11. Questions and corresponding findings after the application of content analysis.

Table 12: ANOVA results for the ordinal Likert-scale questions. The independent variable considered three conditions: variation, control, and non-players (respectively df = 2). Change in students' indications to the Likert-scale questions were the dependent variables for each of the 7 tests.

Question	F	p
Q4i. Anyone can become a leader.	2.61	0.0857
Q4ii. Leaders either focus on tasks or people.	0.77	0.4709
Q4iii. The same leadership principles apply in all cultures.	0.02	0.9793
Q4iv. Leadership is about getting people to do the right tasks.	1.06	0.3561
Q4v. Leadership is about gaining influence over people.	0.13	0.8777
Q4vi. Leadership is about involving others in idea generation and decision-making processes.	0.72	0.4906
Q4vii. Leadership is about empathy and objectivity.	0.07	0.9324

sentences. Content analysis of Q2 showed no observable tendency or other findings. Probably the only phenomenon worth mentioning is that although in the problem statement it was not specified whether the case is in an academic or industrial context, many students assumed that it was academic and thus considered that they were not given the authority to fire the underperforming colleague.

Figure 11b shows change from the pre- to the post-test session in mentions of each stakeholder type in the responses to Q3. It suggests that those students who experienced variation were more inclined in the post-test session to mention customers and less inclined to mention the team. Students in the non player and control groups showed the reverse trend: they were less likely in the post-test session to mention the customer and more likely to mention the team. Both in the experimental and non player group there were fewer mentions of the creditor in the post-test session.

Responses to Q3, on the other hand, lacked the expected focus on the urgent consequence of the occurred risk event – dealing with the creditor. Only 19% of students mentioned the creditor at all, none of them both in the beginning and end tests. In the depth interviews, one student mentioned that she had assumed that there were other sources of funding, so paying the creditor was not an issue at all. This could lead to an interpretation that the wording of the question did not convey the message of the seriousness of these financial implications.

Of the 47 students that completed the pre- and post- test written assessments, 3 did not provide complete responses to all the Likert-scale questions. Thus, 44 remained in the analysis. This resulted in reduced group sizes – for the experimental: 15 participants,

control: 15 participants, and non-playing: 14. A series of one-way analysis of variances found no statistically significant effects for these questions (the F and P values are shown in Table 12).

5.3.4 Role Play Scenarios

Participation in the role-playing assessment was much lower than initially expected. Only 16 out of the 60 students chose to participate. During depth interviews, they provided different reasons for that, including lack of time to engage in an extracurricular activity, lack of perceived value in a role-play for those who already had some professional experience. With some amount of speculation, two more could be added: fear of the unknown and decreased attractiveness, in association with the decreased attractiveness of *vLeader*. One recording out of the 16 was damaged and thus removed from the analysis.

Of the 15 analysed recordings, four were from the experimental condition, ten – from the control condition and one did not play the game at all. Among them some were more successful in engaging than others. Reactions ranged from a student responding to the introductory e-mail in order to calm things down to another student that barely interacted beyond greeting the actors when she entered the room of the role-play. Such behaviours were generally consistent throughout the whole 15-minutes period of the role play, and were thus simplified to single codes. Given this variety, actors managed to do a very good job unravelling the plot so more passive students still get provided the opportunities to engage that emerged at points further down the conversation.

Even among those students considered to have handled the situation quite effectively, there were a number of behaviours that would widely be considered as unproductive in the given context. This involved a number of students overusing various unnecessary gestures (like scratching head, touching face) and *filler sounds* (like uh, er, um). The first few attempts of one particular student to join in the conversation were perceived by actors as quite aggressive and resulted in a prolonged period of subtle but heightened tension in the conversation. At the end of his role-play this student spent some time trying to persuade the actors that the decision he had taken on his own was the right one, but arguably he had only limited success in doing that.

The relatively high number of students that engaged in the conversation only minimally led us to believe that although they were pretty familiar with leadership theories, they experienced other obstacles in applying what they have learned. These might had to do

with language barriers or their personalities. Yet, all foreign students had successfully passed English language proficiency tests and none of them had any noticeable communication obstacles.

Also, throughout the role-play students upon rare occasions used utterances that could be considered beneficial to the conversation, but did not match the framework. Such are cases when students were seeking reconfirmation from the group, like "*Does that sound acceptable for you?*" (P14), "*Are you satisfied with the results?*" (P35). Others were cases of general encouragement, like "*It's not really a crisis. We already just proved in the past that we can solve such tiny problems*." (P35), "So, do you guys have any ideas and stuff?" (P47), "*What are our options?*" (P29). A third set of cases were when participants summarised. Examples for this are "*I would say that we have, like, first to call Barbara, what you did, we could ask Sarah, we could hire someone from outside…*" (P19).

All these instances taken together suggest that students missed a crucial element in leadership learning beyond what they had experienced in *vLeader*. As a consequence all of them could not to apply the leadership principles that they showed evidence that they have rather successfully used in the game.

5.3.5 In-depth Interviews

In this sub-section, I report some qualitative results from the semi-structured interviews. During these interviews, students were asked about the different perceptions and approaches when playing the game.

Discussion of Engagement

In the interviews, some students explained why they did not engage with the game as much as expected. Two reasons were identified. Some students said that, initially, they were unsure of the relevance of the simulation game to their final class grades and, to be on the safe side, they engaged with the game actively. However, later on they realised that their game scores would not affect their class grades. This eased the pressure they felt to play the game. A second reason, reported by students, concerned increased responsibilities as the term progressed, leaving less time to play the game. Some students spoke of the shortcomings of technology for addressing interpersonal relations. During the in-depth interviews, students had explained this in terms of the gradual erosion of their motivation.

During the interviews, two different ways of engaging with the game emerged. Some students took a more exploratory approach to playing the game, trying different strategies

to see what would happen. Arguably this allowed them to experience greater variation. Others focused mostly on the scores they could achieve which did not allow them to experience variation in the intended way.

Most students commented that the activity sheets were too long. Possibly related to their approaches to the game, the way students reported approaching the activity sheets fell into two distinct patterns. In the one case, some students (much as was intended when the activity sheets were designed) read the first activity sheet to understand its structure and from then on used the rest of the activity sheets selectively, according to how useful they perceived each of their sections. In the other case, students would (among other things) consider the ideas table, but in particular, goals and hints where variation was embedded. One student reported that he had informed himself about the background story of the scenario from the game, and not from the activity sheet. Several students explained that initially they had written down the answers in the blank sections in their activity sheets that were left for reflections. Then they went on to complain that because there was not a dedicated discussion on each of these, they lost their motivation to work on these sections during subsequent weeks. Discussions during weekly sessions were short and thus did not relate to each reflective question in detail.

In the depth interviews, some other students reported playing the game ad-hoc without having considered the activity sheets at all.

Possibly this approach was related to their attitude to the game. This had a two-fold effect. On one hand it was reducing their awareness of what ideas would contribute most to their business score. Thus it also additionally undermined their final in-game performance. On the other hand it neutralised any differences, related to variation, as introduced within the experimental design.

Approaches to Play

Several very different predispositions among students were observed. There were clear indications of the presence of motivating factors, matching Yee's (2005) *achievers*, *immersers* and *socialisers*. Some students kept emphasising the importance of competition. A student, illustrating this phenomenon in its pure form commented: "*I did look at the scoreboard, because I wanted to get really high marks, but I didn't look at the details of the game… I was looking at the scoreboard so much that it kind of became like those Nintendo Wii games: I just compete to get to the finish line with maximum amount*" (P18), She was also able to see competitiveness as common to the whole group: "*The*

people we have on our class are very competitive. They all want to win, they want to get the highest marks." (P18)

Another student elaborated on his approach towards discovering how the game works: "*you had to try different strategies, practice with different stuff until somebody suggested [the relevant idea]."* (P02) and "*It is just when I tried different approaches… and I was taking notes of what did I do first in each scenario, what kind of mark did I get. For example if I went positive about this idea, what was my total in the end. Then I will try to go a bit less positive about this idea."* (P02).

vLeader did not provide any socialisation opportunities and students did not take advantage of the discussion space, accompanying the class via Moodle. Still, some students who were inclined to socialise expressed this preference throughout the interviews. For example one student on several occasions demonstrated how valuable socialisation is for her: "*I try to, really, I put it on Facebook or something."* (P55), "*Just don't make me feel like I am useless."* (P55), "*I don't want anyone [to] feel conflict when we work [for] just one month. Just because of one assignment to ruin our relationship, I don't want to do this."* (P55), "*In a group [I need] at least one person that supports me and who wants to work with me. Not [that] everyone just ignores me. Because that gives me confidence."* (P55).

These students clearly expressed motivational preferences that correspond to Yee's types of motivational factors. At the same time the standardised questionnaire that Yee developed to measure such preferences heavily relies on gamer jargon, unfamiliar to randomly recruited study participants or students without particular gaming preferences. This did not allow me to explore in greater specificity the exact relationships between specific motivational drives and learning or application outcomes, even though I attempted to explore this as a follow-up of this study.

vLeader

Several students expressed dissatisfaction of the fact that *vLeader* runs only under Windows. This prevented them from playing the game on their private computers and subsequently had impact on the time they spent playing.

Some students commented that they did not observe any noticeable differences stemming from where they clicked along the continuum of the red-green opinion sliders. Although positive, negative and neutral opinions clearly had different effects on the conversation it

did not actually seem to matter whether they clicked close to the middle of the green (positive) zone, or at its extreme.

To some students the exact utterances, said in result to an interaction act did not seem to be efficient and they felt the need to choose the wording themselves. Some students were frustrated by the fact that when they communicated their intentions, this resulted in an utterance that was very different from what they had been thinking of. One of the students explained it in his own words: "*the voice, the recordings were all predefined and you didn't have much control on them other than clicking right and left. So you didn't have influence to say your own words, to express your own real ideas, and in your own wording, which has a lot of power in a meeting. Two people may want to give positive impressions to the other party, but the words play a big role.*" (P03)

Many students expressed seemingly opposing opinions of the usefulness of *vLeader*. Some of them were very enthusiastic about its learning value, others considered it less useful. These two opinions were combined by a student stating that "*vLeader was OK… Was it useful? For me particularly not, because I've had real life experience before. So for me not much… Obviously it is just a computer simulation, so you can't factor in everything in a computer simulation. But I can certainly see how it would help someone with no or very little experience, dealing with people, managing groups and so on.*" (P03).

5.4 Discussion

The discussion to this study is separated in two parts: substantive and methodological. The first part is intended to discuss what conclusions can be drawn from the results, whereas the second reviews the process used, and suggests refinements for the subsequent studies.

5.4.1 Substantive

In this study, conducted with *vLeader*, the three research questions defined in Chapter 1 are addressed, yet not all manifestations of the objects of learning were manipulated. Whereas I had no influence on the game as a manifestation of the *intended object of learning*, and implemented variation in the supplementary *activity sheets*. Whereas the experimental design of the study did not address the *lived object of learning* and *enacted object of learning*, observations about these two aspects were also made. Due to the unexpected variance in engagement, the originally intended division in comparison groups design was undermined. Discussions of each of the three research questions follows below.

The first research question is RQ1: How does experiencing variation contribute to

improved understanding of learning objectives. Variation was introduced into the wider intended object of learning via the supplementary materials. Answers to Q1 showed that students were more inclined to describe leadership in terms of influence, which indicates how their way of thinking of the subject has come closer to that of researchers in the field (Yukl, 2005). Although this was a common trend between all the groups, it was stronger with those playing the game and even stronger in the variation group, which indicates that the game and variation could have helped them to reconfirm what they've been taught in the class.

Although Q2 did not lead to any generalisable findings, it gave students an opportunity to solve a problem that is very close to their personal context. This effect was reconfirmed by the fact that some students assumed further unstated contextual settings, similar to their own environment.

The results regarding awareness of a particular stakeholder in Q3 could be attributed to experiencing variation in scenarios 3, 4 and 5. The first two of these scenarios emphasised the role of others in the decision-making process whereas Scenario 5 focused on awareness about the three business aspects: financial performance, customer satisfaction and employee morale.

A probable contributing factor towards the lack of statistically significant differences in the Likert-scale questions is the fact that the questions were not related to particular scenario and the corresponding learning objectives, but were more general questions about leadership. Thus these fixed response questions were unable to clearly separate learning in class from the added value of the game.

Although the intended separation through the learning materials did not take place as planned, during depth interviews some students expressed preference towards learning in different ways. Some of these showed predisposition towards experiencing variation, thus affecting their *lived object of learning*.

Meta reviews show that absence of strong quantitative evidence of learning is common in the areas of problem-based learning (Gijbels et al., 2005) and instructional games (Hays, 2005). However, the weak evidence quoted above show two indications. First, the use of the game was beneficial in the learning process, although benefits are spread unevenly across participants. This imbalance could not be attributed only to level of engagement, but was probably influenced by some other factors that were left unobserved. The second

indication is that although the presence of noise (in the form of people not following procedure) needs to be acknowledge, there is a tendency that the introduction of variation in the content of the learning materials (which is the intended object of learning) was beneficial to learners. Further studies need to examine what type of learners benefit from game-play and variation, respectively. This might be connected to the discussion of the effect of variation on engagement.

In order to address RQ2: How does experiencing variation contribute to improved application of learning objectives, a practical assessment procedure had to be introduced. This was done through the role-play. The cases of students' shortcomings in the role-play indicate that computer-mediated environments could possibly allow for only limited practice which could possibly have to do with shortcomings in realism of the simulation. In this particular case, student did not get to practice the effects of real communication between people. Aspects such as control over own non-verbal communication and pressure in face-to-face communication were not practised at all. As a consequence, some students were noticeably unprepared for a role-play assessment. It could be speculated that similar behavioural limitations could have been experienced in a more realistic situation.

Apparently there is need for further research in how certain desirable behaviours could be encouraged. The most promising currently identified approach is through practice, but there is little consensus on the specific form of leadership practice. Part of the questions around this practice have to do with whether these should be computer-supported experiences at all.

Notwithstanding that it is crucial and open question, I consider the question of how the practice of computer-based serious games can be made to replicate an environment where learners would exhibit real-world behaviour to be beyond the scope of this research thesis. That's why I consider it infeasible to reach a convergence between results of learning with serious games and simulation of real-world situations. That's why, for other studies in this thesis, I limit tests for the application of newly developed knowledge, to application within the context of the serious game itself – this being an instance of problem-based learning (PBL)

The third research question that was originally considered is RQ3: How does experiencing variation in the learning content (in other words, the *intended object of learning*) contribute to higher engagement. The variation group engaged more with Scenario 1 (see Figure 9 in Section 5.3.1), but not with the others. The insignificant difference after the

first scenario could be attributed to the gradual loss of interest on one hand, and to the less attention students paid to activity sheets (where variation was actually embedded). As a consequence, Scenario 1 provides a positive indication for RQ3, but further evidence is needed.

Something else partially related to the findings about RQ1 is that during the depth interviews some indications related to variation have been detected. These had to do with how students actually experience it. In contrast to traditional learning, taught by a teacher, with games for learning players have more autonomy in how they learn. In order to make them experience variation, the game design should accommodate for their own predispositions (*lived object of learning*) and embed variation in the experiences they are likely to have (*enacted object of learning*). From exploring player preferences in depth interviews, it appears that personalities of participants could be loosely matched to player preferences, as they were identified by Nicholas Yee (2005): *achievers*, *socialisers* and *immersers*. Whereas the current study allowed people with preferences leaning to the third group of immersers to experience variation, others that have given indications that they are rather achievers or socialisers, did not engage equally. This seems to an indication that learners more inclined to experience variation may be better at learning with games.

5.4.2 Methodological

In order to encourage students to get engaged, two possible options have been identified. One of these options is to employ a within subjects design of the study with an intermediate assessment. This would both overcome any regulatory and ethical concerns of providing equal access and allow for separation of the effects of each of the experimental conditions. Furthermore, it has the potential to allow exploring the effects of different modes of delivery over time. The other option regards identifying other universally attractive incentives beyond money that would allow students to engage willingly. A possible candidate for such an incentive in a university context could be additional degree credits. Yet, this would introduce further complexities as providing a hypothetical advantage to the experimental group.

Also, introducing variation through auxiliary materials involves the risk of students ignoring these materials, and thus not experiencing the intended variation at all. As a result a recommendation could be made that in future studies variation should be embedded directly in the game, so it is not so straightforward to find a way around it.

A number of other indicative findings from this study could be considered as source of other useful feedback. The three main outcomes could be considered relevant are the excess length of activity sheets, doubts in reliability of system-collected performance data and limitations of computer-mediated learning environments.

According to student feedback, activity sheets need to be made shorter, preferably restricted to one page. The sections about learning objectives, scenario background and context and learning refection could be removed. In line with Biggs's idea (Biggs and Tang, 2007) that whereas teachers focus on intended learning outcomes, student focus on assessment, evidence from this study indicates that activity sheets do not need to contain intended learning outcomes. Students reported that they have ignored them.

As explained in Section 6.3.5, where in-depth interviews are analysed, scenario context and background in the activity sheets was also not generally considered useful by students. It could be speculated that in the context of an interactive multimedia experience, such as *vLeader*, providing background information in the form of text is not attractive enough and does not get students involved with it. Different ways should be sought to deliver this information in a way that is more coherent with the overall multimedia experience, possibly more similar to what was done for the role-play.

As for the success planning and reflective questions, despite the widely acknowledged value of learning reflection, they were not useful in their current form. There are indications that students were not satisfied with just putting their answers down on paper, but also needed some form of feedback on these answers. Possible improvements could focus on separating them from the activity sheets and relating them closer to class discussions. A further relevant question is whether these discussions could have been more successful if an attempt was made to facilitate them through online forums. This failed to happen using UCL's Moodle system, although both the lecturer and researcher attempted to engage students by maintaining frequent updates of information through its forums.

There are certain issues with performance data, collected by the game. Two of them are worth mentioning: the concept of separation of *practice*/learning and *advance*/assessment modes is meaningful only when players have a clear understanding of how assessment is being formed and such performance data is prone to a number of distortions and thus its validity could be questioned.

Chapter 6: Using Rapid Prototyping to Develop a Serious Game

6.1 Background

Developing a game using rapid prototyping means having two concurrent and interacting streams: development (a case of *design-based research*, progressing from paper-based materials to software) and evaluation. These correspond to the materials and procedure of the evaluation studies and are described in dedicated sections below. In this chapter I report the development of a learning game and early studies conducted with it. The development took advantage of rapid prototyping and frequent formative evaluations. The initial requirements were a preliminary set of intended learning outcomes and a theoretical framework. I used an established framework from conventional crime prevention called Conjunction of Criminal Opportunity and abbreviated as CCO (Ekblom, 2011a), and applied it to information security, as others have previously suggested (Collins and Mansell, 2004). The development was intended to make both the framework and the learning outcomes more widely accessible and to reduce the facilitator effort needed to teach them.

I conducted five small-scale formative evaluations to generate user requirements and feedback. These involved a total of 7 security experts, 6 usability experts and 7 university students. Once the software has reached sufficient maturity, further learning studies were conducted with 28 randomly recruited participants (reported in Chapter 7) and two consecutive class studies with feedback collected from 23 participants (Chapter 8).

This work has been reported in several dedicated publications for the concept (Ruskov et al., 2014), the development (Ruskov et al., 2012) and the learning assessment results of the lab evaluation (Ruskov et al., 2013). These and other (unpublished) results are explained and discussed here.

Research in the recent decade (Adams and Sasse, 1999; Stajano and Wilson, 2011), demonstrates the importance of addressing the human factor in information security: attackers often obtain information, access to systems or money by tricking customers or employees. This calls for security models that capture the complexity of the wider socio-technical system. Examples of such models are the Mechanics of Trust (Riegelsberger et al., 2005) and the Compliance Budget (Beautement et al., 2008). Of interest is the unification of such models in the search for a more general theory of how we can design

systems that prevent this sort of attacks and consider the multitude of relevant factors.

Another related field of research – crime prevention in the material world – has accumulated more experience, in this regard, than information security. Research here, under the discipline of crime science has included social aspects when considering both personal and situational factors that play a role in crime. While most approaches in crime science have focused on the crime situation, more holistic perspectives which integrate this with a richer understanding of the offender have been claimed to be advantageous (Ekblom, 2011a, 2007).

The consideration of these resulting combinations leads to an increase in complexity which, although necessary for getting to grips with real-world crime, makes the problematic considerably harder to grasp. This is even more challenging when the framework needs to be brought to the attention of employees who are not professional security staff. Such people have their primary business tasks and despite its importance, security is secondary to them (Albrechtsen and Hovden, 2009). Yet, if people get more efficient in understanding the role of security, they will see the need and understand the value of applying it. In other words they will also become more adept at contributing to solutions as intelligent users.

The Collins and Mansell (2004) report mentioned previously and prepared for the UK Government Foresight Programme, suggests the adoption of CCO framework as a basis for developing and designing resilient systems and effective cyber defences. The CCO framework was developed as a generalisation resulting from the analysis of several thousand crime prevention projects implemented through the UK Safer Cities Programme (Ekblom, 2011a). Its typical original application was the analysis of criminal hotspots wherein certain types of crime recur in a small vicinity.

This framework presents a systematic and conceptually rigorous categorisation of immediate contributory causes to criminal events (with the potential to trace back to distal causes). Compared with prior frameworks in crime prevention (Bullock and Tilley, 2011; Clarke and Eck, 2009; Knutsson and Clarke, 2006) it captures a much wider range of causes for attacks. CCO comes at the price of greater but necessary complexity relative to frameworks that are currently widely used. Considering each of the eleven generic causes visible in the CCO diagram (see Figure 12) equips practitioners to better handle the complex reality of crime (Ekblom, 2011a). Identifying potential causes naturally leads to ideas for their intervention counterparts. These intervention ideas could block, weaken or

divert the causes, such that the criminal event is less likely to be attempted, or to succeed. For example an attempt to reduce the absence of crime preventers could be the introduction of security staff to undertake surveillance; an example for securing the enclosure could be the introduction of access control; and the anticipated risk, effort and reward could be reduced by means of deterrence like the introduction of penalties.

CCO was designed to contribute to the *preventive process*, a very basic outline of which is set out in the following steps (Ekblom, 2001):

- Identification of crime problem – the symptoms – and setting of objectives for reduction
- Diagnosis of causes of crime problem
- Selection of specific interventions, and creation of practical operational solutions
- Implementation

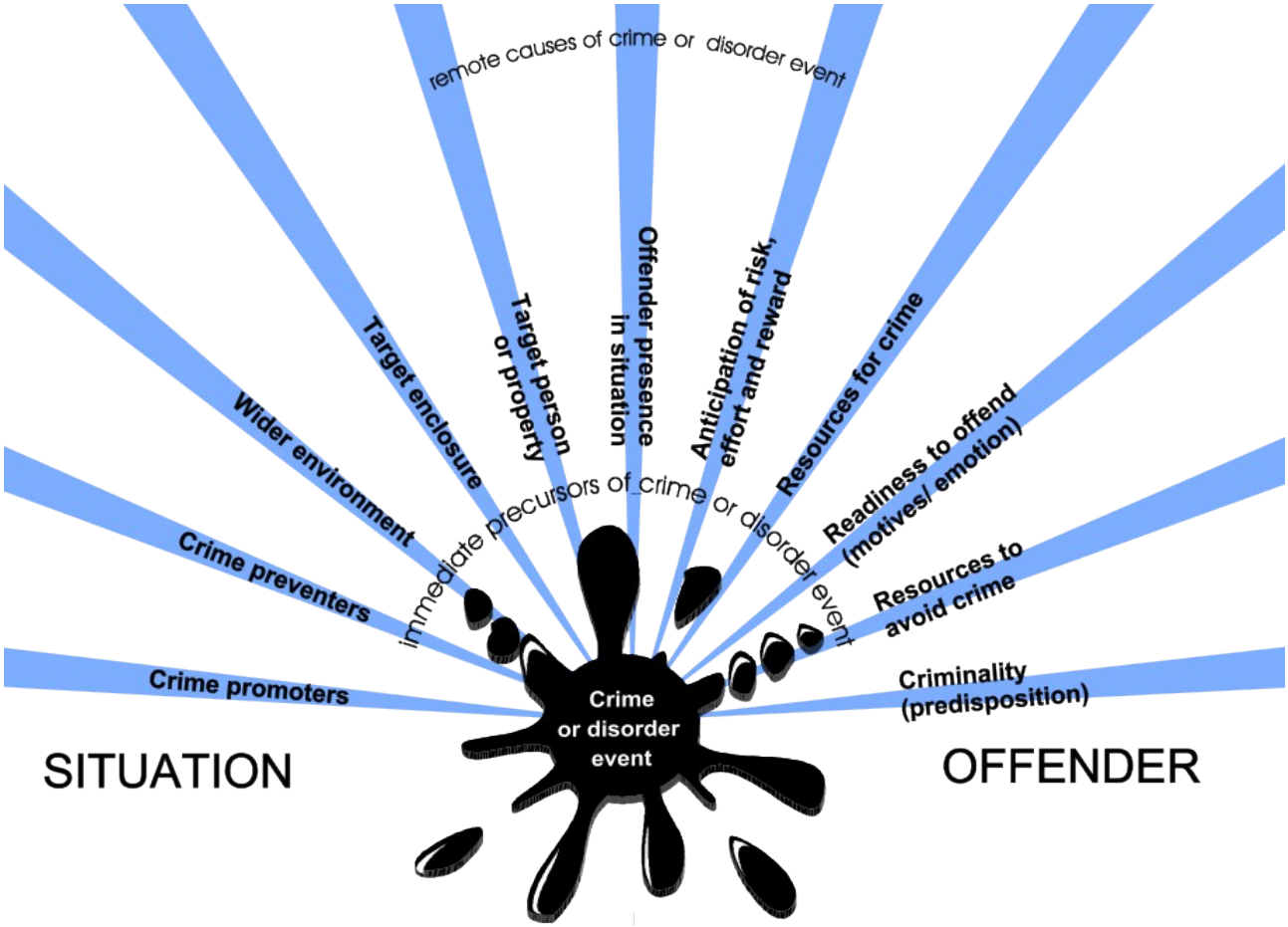


Figure 12: The Conjunction of Criminal Opportunity (CCO) diagram as presented in (Ekblom 2001). It features the eleven generic causes that when combined give rise to the criminal event. On the left-hand side of the diagram are situational factors, and on the right-hand side are factors from the perspective of the potential attacker (i.e. offender).

- Evaluation and adjustment

The variety of possible intervention ideas and the exact details of their implementation lead to a classification of how specific, or general these ideas are. The CCO framework distinguishes between *principles* and *methods*. Methods are the detailed practical activities that are necessary to be conducted to implement the intervention. Research shows that these are often difficult to transfer to other situations – what works in crime prevention is very context-dependent (Tilley, 1993). Principles, on the other hand, are the more abstract description of what is being done that is formulated in a way that could be re-applied, customised to context in other situations. Let's consider the possibility of employee satisfaction probes, which not only collect information, but also allow employees to "let the steam out". The principle behind them could be the need to allow employees to talk about problems, and then if necessary address, employee attitudes for disgruntlement (as an instance of *'readiness to offend'*). A common implementation method involves collecting satisfaction feedback at annual appraisal interviews, and of course acting on the results.

This makes CCO generally applicable, yet simple enough to serve as the theoretical foundation of a simple serious game for the purposes of this research. Moreover, challenges encountered during the development and exploitation provide feedback that tests the consistency of the theoretical research, as has been done with simulations in conventional crime science by Birks, Townsley and Stewart (2012).

Although crime prevention practitioners report that they recognise the usefulness of CCO, more often than not they also find it challenging to use, because of its complexity and subsequent difficulty to oversee all the systemic interactions that it represents. The aim of the software – with working name CCO prototype – was to engage players driven by different motivational factors, and to get them to experience the variation necessary to effectively learn how to use the framework. Given the complexity of the modelled theory, the prototype design needed to maintain perceived simplicity in order not to unnecessarily confuse players any further.

A discussion of the details of CCO and how it could be adapted for the purposes of information security follows. Three theoretical models of insider attacks are presented here and their corresponding representation in CCO is discussed.

In his research on insider attacks Schultz (2002) considers the CMO model consisting of *capability* to commit attack, *motive* to do so and *opportunity* to do it. In his work Schultz

also reviews a model of insider attackers by Tuglular and Spafford (1997) allegedly suggesting factors such as *personal characteristics*, *motivation*, *knowledge*, *abilities*, *rights and obligations*, *authority and responsibility* within the organisation, and factors related to *group support*. Parker (1998) develops the SCRAM model. The abbreviation corresponds to the considered factors: *skills*, *knowledge*, *resources*, *authority* and *motives*.

The CCO framework considers eleven overarching classes of complementary causes which come together to make a criminal opportunity. The number of these classes may seem high, but that is to model all potential contributory causes of crime in the real world. To model the contributory causes of cyber crime, the classes of causes can be considered to fit in three wider groups – personal, technical and social factors. These all are represented in Figure 12.

The personal (attacker) factors are:

- Criminality and wider personal traits influencing attacker's tolerance towards immoral or criminal deeds. This is where personal characteristics (Tuglular and Spafford, 1997) are being addressed.
- Anticipation of risk, effort and reward – the rational decision and utility behind the attack
- Abundance of resources to commit crime – both cognitive resources and capabilities, and social factors such as trust, but also technical hacking tools – the attacker needs to be both aware of their existence and be able to operate them. Schultz's (2002) capability and opportunity can be viewed as part of the cognitive resources of inside attackers. Parker's (1998) resources fall naturally into this category, but also his skills, knowledge and to some extent authority (refer to wider environment for another cause that corresponds to aspects of it).
- Immediate readiness to undertake an attack, for example the commonly modelled emotional/motivational state to do it like disgruntlement (Parker, 1998; Schultz, 2002).
- Lack of skills to avoid crime – potential skills that would reduce attacker's need to commit crime. For insiders this could be ability to manage stress, soft skills to improve common understanding of potentially discouraging issues, etc; ability to

secure legitimate advancement in current company or beyond.

- Attacker presence in situation – circumstances like the fact that this person is part of the organisation, but also that they have certain access privileges that might allow them to abuse the organisation.

Situational factors contributing to the criminal opportunity capture the potential (or possibly *affordance*) of the environment to allow for criminal behaviour. In the context of cybercrime as such can be seen aspects like scamming and social engineering (Stajano and Wilson, 2011). These could be both social and technical and are represented by the following:

- **Presence of crime promoters** – people (or roles, or organisations) that either deliberately or unintentionally make the crime more likely, such as potential buyer of stolen data, hacker technology providers or someone who carelessly does not log out from the system after having used it.
- **Absence of crime preventers** – people that intentionally or not would discourage, deter, block or reform attackers, be they security officers, technology providers or management and staff ready to help disgruntled colleagues resolve their grudges.
- The **target** of the attack, in the digital context commonly information or finances, but could also be the person or organisation which might suffer from a leak.
- The **enclosure around the target**, having both the digital aspect of firewalls and authentication systems and physical access.
- The **wider environment** which could contribute motivationally or tactically towards attacks and discourage or restrict potential preventers. *Authority* can be thought of as a feature of the organisational environment that allows misuse, like lifting doubt from people with authority that demand information which they don't possess.

The issue of *authority*, as considered by previous models, comes to illustrate the interactions of these different factors. On one hand *authority* is a *resource* that a potential offender possesses and enables them to commit the attack. On the other one's *authority* also is a function of the *wider environment* – it is the organisational culture that allows authority to be exercised without protective questioning from others.

6.2 Toolkit Design

The challenge of designing effective serious games actually addresses to a factor, identified to contribute towards information security compliance behaviour, as identified by Pahnila and colleagues (2007). This factor is *information quality* as perceived by employees and in the author's study it was shown to influence actual information security compliance.

A consideration of de Freitas and Oliver's (2006) evaluation framework is useful in shaping the future design. The four dimensions of the framework (context, learner specification, pedagogic considerations, mode of representation) are considered here.

Similarly to *vLeader* and many of the games considered in Chapter 3, the *CCO toolkit* is intended to be used by students in a class as an additional self-paced activity. Given that the toolkit is web-based (and thus possibly accessible to anyone), this is taking advantage of the fact that class attendees typically have a more homogeneous background and are motivated to explore the topic at least to the level to sign up for a class.

Due to practical constraints of the research environment, the toolkit and accompanying materials are aimed towards the audience of students in a class called People & Security. The toolkit was made with the intention to engage students in solving real-world problems, and provide as insightful and diverse feedback, as possible given that implementing the actual solutions is not within the learners' reach. Different forms of group work were explored (see both pilot studies with paper prototypes described in Section 6.3 and the first class study in Chapter 8).

In contrast to the common trend outlined by Kebritchi and Hirumi (2008), and to a huge extent in response to it, this development employs Davies and Mangan's (2006) theory-principles-activities approach and develops it further in the model seen in Figure 2. As part of the design, intended learning outcomes, relevant for the considered classes, are formulated (see Table 13) and are embedded in the design. The actual learning activities are the familiarisation with the terminology and paradigm of CCO, the identification of ideas for causes and interventions, the evaluation of ideas of others, and the review of feedback received by others. In line with the identified good practice, briefing and debriefing sessions are also planned and designed.

Even if the mode of representation is an important factor in learning, due to the limited resources for the research conducted within this thesis, it was decided that a gamified prototype toolkit will be developed, instead of a full-blown serious game. This would

Code	*Intended Learning Outcome (ILO)*	*Type of Knowledge*
ILO1	Understand what exactly CCO is, what it's for, and the wider process in which it can be used.	Declarative
ILO2	Use CCO to interpret causes of criminal events within the worked examples.	Procedural
ILO3	Use CCO to identify preventive intervention principles that they could bring to bear against these causes.	Conditional
ILO4	Generate greater numbers of plausible intervention ideas – i.e. the first stages of innovation.	Conditional
ILO5	Grasp of the key concepts, e.g. ecological level.	Declarative
ILO6	Use CCO terminology correctly.	Declarative

Table 13: The intended learning outcomes (ILO) that were considered to be important to achieve with the game-based toolkit, and the corresponding type of knowledge according to Sugrue's taxonomy.

certainly undermine the immersion, as it has been evidenced in Study 1, but an attempt to mitigate this is including interview questions involving usability and immersion, so that recommendations can be made for the future development into a serious game.

Coming back to the pedagogical theory, and as suggested in literature (Biggs and Tang, 2007; Davies and Mangan, 2006) learning needs to be driven by clearly specified intended learning outcomes (ILOs). Six such intended learning outcomes were sought when learning the CCO framework with the toolkit. These were listed and classified according to Sugrue's taxonomy of knowledge as illustrated in Table 13.

Although identified in Study 1 (see Section 5.2) as a factor in engagement to play and performance in the game, the introduction of distinct playing modes for learning and performance was considered to be of lower priority than other features specific to the learning goal and experimental setup.

Initial brainstorming to provide the four necessary conditions of learning (Marton and Pang, 2006) resulted in the following way to explicitly implement all four: *contrast*, *separation*, *generalisation*, *fusion*. For the early prototype, all four are expressed through role-play in the peer assessment. This means that after generating their own ideas, players are asked to assume various roles from the scenario, are given ideas suggested by others and are asked to provide feedback. Seeing new realistic ideas covering the full spectrum of CCO exposes players to *contrast* and *separation*. While role-playing, players provide feedback from the perspective of an offender or potential victim. This allows them to maintain focus on CCO while considering the context of different preventive interventions and perspectives, thus experiencing *generalisation*. The prototype allows players to review their own ideas once again at the end and thus *fuse* between their initial ideas and what

they have seen during role-play.

A number of lessons from Study 1 and previous studies (Bocconi et al., 2012; Malheiros et al., 2011; Ruskov et al., 2010) were utilised in this development. Several examples of how they were reused for the *CCO toolkit* prototype follow.

A key lesson from an earlier study (Ruskov et al., 2010) was that design needs to help users, driven by different motivational factors, to converge toward the intended objective. In the interviews conducted after Study 1, people engaged by Yee's motivational factors were also identifiable. Different users must be guided through the necessary experience to effectively learn how to use the framework and generate and discuss crime prevention ideas. This meant that in the design of the *CCO toolkit*, different features needed to be designed to attract types of players who were driven by different factors.

A first approach within the rapid prototyping process involved applying Yee's motivational factors (2005; 2012). This resulted in addressing each of the three wider categories: *achievement*, *social* and *immersion motivations*. *Achievement* is the drive to stand out and excel; it is met in the prototype by introducing a scoring and ranking system, based on peer assessment of user ideas, and dedicated badges for suggestions that are perceived as innovative or described with exceptional detail. The need for meaningful interaction with other people – the *social motivational factor* – is addressed through the opportunity to comment on ideas. It is also intended that the prototype is also usable in groups, thus fostering face to face discussions. The *immersion factor* – getting engaged in the process of the game – is addressed by the problem-based formulation of the task, and the exploratory breadth provided by the variety of 11 CCO generic causes and counterpart intervention principles.

For example scoring mechanisms needed to be put in place to engage achievers. This could be compared this with the multitude of players in the *We The Giants* study (Ruskov et al., 2010) that focused on the game world and mechanics, and on talking about the task of reaching the star. The *CCO toolkit* also needed to provide opportunities for discussion to engage players socially, as this need was identified by meta-game references as captured in that prior study.

Furthermore the multitude of commonly recurring pieces of advice (be it talking about the example, the game world or the giants theme) potentially allowed for an attempt to automatically cluster players' short-text contributions. This is possible when the game

guides players to converge their contributions towards the objective of the game. As a result, words were commonly used in a narrower context than in general-purpose conversation and domain-specific language was being used. On the other hand, designing and implementing a simulation of information security behaviour is a challenging task. A computer simulation typically requires a finite (and thus mathematically closed) representation. Considering that there is an arms race between attackers and security officers which requires continuous adaptation and innovation (Ekblom, 1997) – a potentially infinite space of ideas or steps. A way out of this contradiction could be to address only the recurring, 'bread-and-butter' attack and security ideas, but not the innovative ones. This way the hope would be to get coverage of the "20% of scripts occurring 80% of the time". Contributions could be automatically classified to capture repetitive ideas and cluster them in commonly recurring classes. This idea matching mechanism could allow for immediate feedback when players share, for reuse of assessment and for comments on previous similar contributions and higher chances of identification of innovative ones – those that do not match any existing class.

6.2.1 Software Development

A widely used form of rapid prototyping in software development is a set of techniques collectively called agile methodologies. An adaptation is used for the purposes of prototype development. Loosely based on *Scrum* (Kniberg, 2007) the development effort consists of 2-week sprints. In effect overall goals were identified, but priorities of tasks to be implemented were planned only for the following sprint. Tasks need to be broken down to a complexity that could be implemented in just a few days. Then the effort needed to implement them is estimated by the developer. As a result of this process there is a growing list of tasks that have been identified as potential improvements, but are considered of too low priority to be implemented in foreseeable future.

To facilitate this process a free online software project management platform is used. The features that are intensively used are version control, ticketing system and milestone planning. It also provides an RSS feed for any activity making work on the project easy to follow.

In order to allow for browser use (subsequently also touchscreen devices) and flexibility well-established open source web technologies were chosen for the development platform. HTML/SVG and JavaScript/JQuery are used for the client side and PHP as web services with MySQL are used on the server-side. It was estimated that the technologies of choice

are sufficient for the early requirements of the browser-based prototype and individual technologies can be replaced with little effort. As a way to focus the work on the prototype, a decision was made that only the Google Chrome browser will be supported – a platform that provides an optimal balance of ease of development and popularity.

The employed technology allows for the creation of a burnout chart – an open-ended variant of the burndown chart as used in Kniberg (2007). The chart measures days of development over time. It illustrates the long-term and short-term scope of the project on one hand and the expected and actual progress rate on the other. The long-term scope is defined as tasks that are identified and short-term scope is the subset of tasks that are considered realistic for the next version of the prototype. They increase throughout the project, because new tasks are identified either because of more detailed planning or as a result of user feedback. Expected progress is one abstract unit of development for every working day. Actual progress is measured by the rate of marking tasks as complete on the ticketing system.

6.2.2 Prototypes

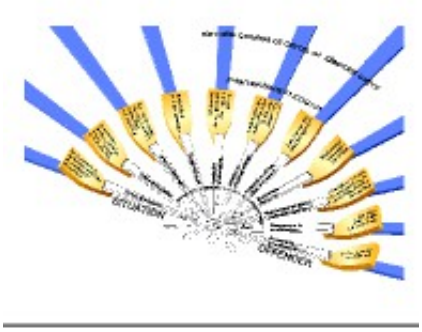
My approach to rapid prototyping included early paper-based versions of the intended game interaction and, once development moves to software, adapting the *Scrum* agile development methodology (Kniberg, 2007). In response to formative feedback the development process included two paper prototypes and a subsequently and continuously developed software prototype with stable releases every two weeks. The narrative of this scenario also evolved in parallel with the development of the prototype. Before the *Meltdown* scenario (see Appendix C) was developed, prototypes were based on two scenarios that each featured a recurring conventional crime/disorder event. By using each of the prototypes users are asked to come up with contextualised cause and intervention ideas, and subsequently to assess such ideas, proposed by others.

The development of work-in-progress prototypes is described here. The final version of the web toolkit, used in the studies is described in a subsequent section of this chapter (see subsection on web-based toolkit).

Paper prototypes

The first paper prototype was developed with the intention to encourage participants to both decide how they could re-use existing crime prevention good practices for the scenario, and also be innovative and come up with their own ideas. I designed a board-

and-cards prototype to be used by a group. In it the discussed scenario was presented through a narrative text, a map, and a set of photographs of the environment. When using the prototype each participant received a board with the CCO radial diagram (see Figure 12) and a stack of 50 cards with ideas on them (an example provided in Figure 13). Participants were also given blank cards and a pen to suggest new ideas whenever appropriate.



Take account of lighting, CCTV, surveillability and scope for response by guards, in integrated approach to improve visibility in remote areas of the car park.

Figure 13: A card with a pre-suggested idea, representing possible intervention method for a conventional crime prevention scenario

After the board-and-cards prototype, a second pen-and-paper (utilising carbon paper) prototype was developed. This second prototype used the same board as the first one. However, instead of providing participants with ideas in the form of cards, it allowed them to freely write ideas on the board. They were offered only one idea as a 'starter' example to refer to. A facilitator guided participants through a process of getting to know the scenario, 3 phases of generation of ideas and 3 phases of role-play and revision of ideas. At the end the facilitator collected the materials for expert assessment (both roles were typically performed by me, except for the expert assessment reported in Section 7.3.2).

Browser-based prototype

Participant feedback from the second paper prototype was positive (more details in the procedure and results sections) and it was chosen as a basis of a computer-based version. For this, a website with a Model-View-Controller architecture was developed. It is discussed in the rest of this chapter. It features the same process, plus a final score screen for immediate automated feedback when possible. The overall process can be split into scenario, idea generation, role-play assessment and score (providing feedback).

The first screen after a user logs in is the scenario to introduce the learner to the problem at hand, which is initially a scenario about car parts theft from an airport parking lot, then the *Moonshine* scenario (previously developed to teach CCO), and finally the *Meltdown* scenario, which was the only one addressing information security.

After initially taking learners (users of the software) through reading the scenario, in the idea generation part, the browser-based version features a clickable version of the CCO diagram. Whenever a learner clicks on one of the 11 contributing causes in it, further

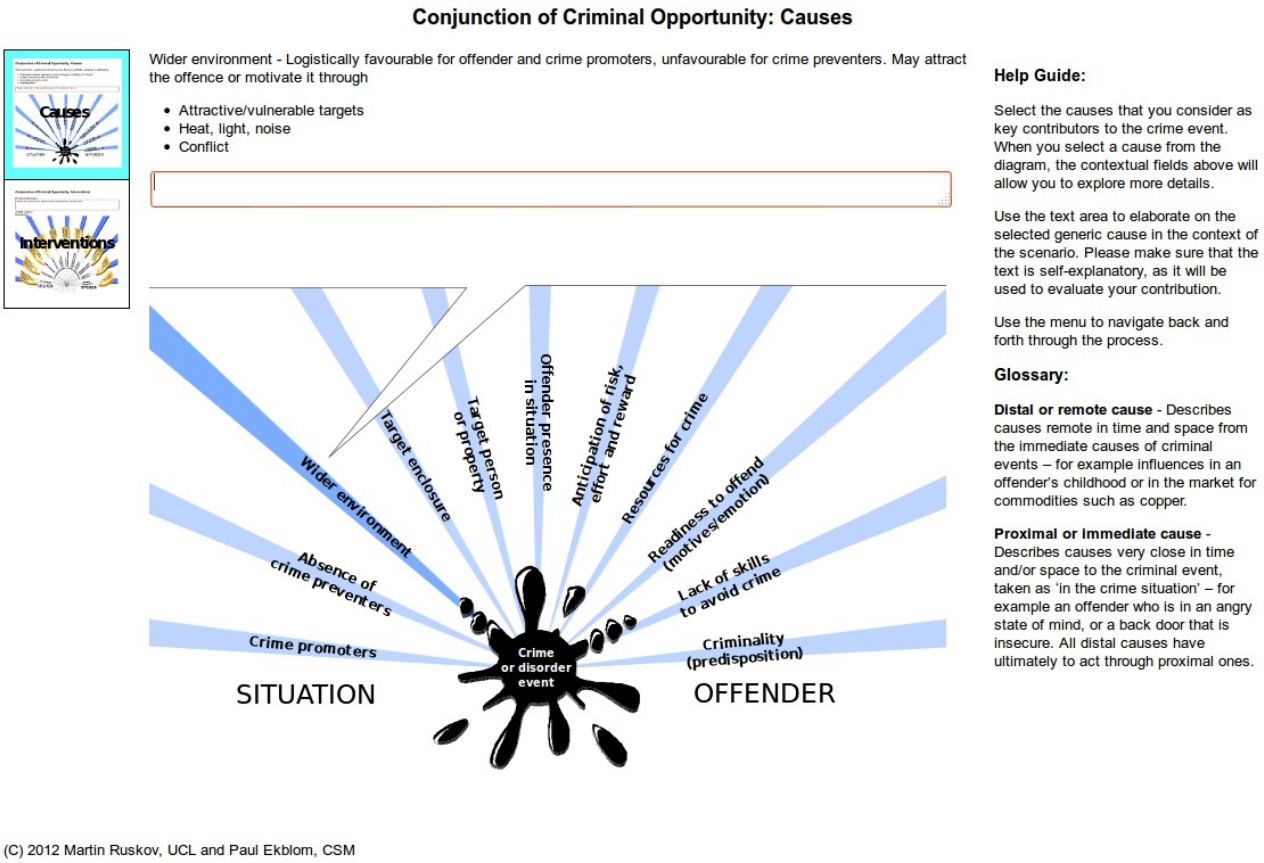


Figure 14: Sample screen of the Interactive part of the browser-based prototype also featuring the CCO diagram used throughout the studies. On the left is a navigation menu allowing participants to switch between currently accessible views. On the right there are static context-specific explanations.

information appears and the learner is prompted to type ideas within a dedicated dialogue box (see Figure 14). Learners are sequentially taken through two related screens with the diagram, one for ideas for causes, and another one for ideas for corresponding interventions. The ideas (short textual description, which is text-based qualitative data) collected in each of the screens are further reused. Ideas for causes are prompted as reminders at the interventions screen and ideas for interventions are worked with further as explained in the remainder of this section.

After the phases of generation of causes and interventions, learners are given access to a table having the proposed interventions as rows and the 11 CCO generic causes as columns. Learners can tick table cells where they think possible side effects to a planned method may occur. This way they can further explore the impact certain intervention could have on a wider set of causes than merely the 'focal' one.

The assessment part of the process prompts learners to evaluate ideas of interventions generated by other learners (for example previous participants or students). This is done by letting learners grade each idea along a 5-point Likert-scale (Hartley and Betts, 2010).

This way qualitative text data is given quantitative assessment by other learners. In the assessment each learner is first asked to assess how strongly each idea impacts the criminal opportunity. After this initial assessment, two more follow, this time asking learners to think from a different role perspective. One of the subsequent assessments puts them in the role of the offender assessing how ideas impact upon the criminal opportunity. In the last assessment learners are put in the role of neutral citizen who could potentially become victim of crime, helpful preventer (someone who by their action or presence makes crime less likely) or even unintentional crime promoter (someone who inadvertently makes the crime more likely to happen, for example by forgetting or neglecting to lock a door). In contrast to the previous assessments screens where probability of criminal events is peer-assessed (thus focused on), in the third one learners are asked to assess if intervention ideas have impacted the harm that the criminal event can cause once it happens.

The final game screen delivers the score achieved by the learner. This includes overall score and ranking of the learner, and a table with suggested intervention ideas and a breakdown of the three scores these ideas received from other learners. In order to be able to deliver feedback instantly at least some partial assessment of ideas is necessary. To this end, newly suggested ideas are matched against patterns derived from the database of previously suggested ideas. When a new idea can be matched to a pattern with good certainty, it is automatically assigned the average assessment of previous ideas matching that pattern.

The Scrum development process generated a growing list of tasks that had been identified as potential improvements, but were considered of too low priority to be implemented in the foreseeable future. To facilitate this process an online software project management platform was used. The features that were intensively used were version control, ticketing system and milestone/*sprint* planning. In order to allow for browser use and flexibility established web technologies were chosen for the development platform, notably JQuery and SVG.

6.3 Procedure

The prototype is designed to address knowledge which requires users that have crime prevention expertise. Researchers from the Information Security group at University College London (UCL) were considered to be subject matter experts (SMEs) and approached with a proposal to be participants in studies with the prototype. One class of

MSc Crime Sciences students was used to test the paper prototype in a class setting. Three small-scale formative evaluations were conducted with the paper prototypes and two with an early version of the browser-based prototype. These are summarised in Table 14 and described in further detail below. For one of the studies it was considered that usability expertise would be useful, so correspondingly usability experts were recruited.

Paper prototype

One evaluation was done with the board-and-cards paper prototype and two with the pen-and-paper prototype. The first took place as a one-hour game-playing session and a subsequent half-hour focus group with four participants (assuming the role of learners). They were given a personal set of board and cards and a pen to write on blank cards. Participants were given an explanation the scenario and asked to discuss and propose ideas by each putting them on their personal board. At 15-minute intervals they were asked to change places and discuss the ideas of others.

The second prototype (pen-an-paper) was used in two different evaluation sessions. The first was intended as a focus group with three participants, but due to absent participants actually turned out to be a one-participant one-hour think-aloud session. The second was a deployment of the prototype in a class environment. A class of 7 students was split into three groups and each group was given a folder containing the carbon paper version of the prototype. All participants engaged with providing ideas intuitively when they were told they could write on the A3-sized paper diagram they were given, representing the game board. In these circumstances participants were given 45 minutes, but that time was insufficient for the role-play assessment part of the process. As a result it had to be omitted from this formative evaluation.

After the successful studies with the second prototype it was decided to move on to a browser-based prototype replicating the paper-based interaction of writing on segments of the diagram.

Browser-based prototype

The browser-based prototype was under continuous scrutiny and discussion by members of the development team, including non-technical subject matter experts (SMEs). So far it has already gone through two evaluation studies with participants. Three usability experts commented on it and two SME participants used it in a think-aloud session. The usability experts provided 50 recommendation items in total. Recurring ones had to do with the

oversight of the process through the navigation menu and distributing information in on-demand pop-up dialogues to reduce its perceived volume.

SME participants took about 45 minutes to complete the game process without being pressured by time. In the think-aloud sessions participants were very positive about the idea of providing a dynamic interface to introduce something as complex as CCO.

6.4 Results

These five small evaluations reconfirmed the usability and learning potential of the prototype. Typical for rapid prototyping, this feedback provided the development with clarity about short-term steps that needed to be taken and some vision of what a future browser game should look like.

Early results from the knowledge sharing mechanism piloted with this prototype are encouraging. In the first paper prototype participants got focused in reading the multitude of ideas presented to them with the cards. On several occasions they were reminded that they could suggest their own ideas, which resulted in discussions, but it actually required input from the facilitator to get the raised ideas written on new cards. In the focus group they commented that they did not find the map and photographs useful. This prototype was considered unsuccessful, because it overloaded learners with options, thus causing *paralysis by analysis*. Once this obstacle was removed and learners found it easier to contribute, they readily completed nearly the whole wide range of contributions, resulting in more than 25 ideas each on average.

SMEs evaluating the browser-based prototype provided suggestions of how to improve the usability and attractiveness of the prototype. They further provided some comments that were very different from those of usability experts, but overall there was considerable overlap between the two SME participants. They felt confused about the terminology used by CCO, as it somewhat differs from theories that participants were more familiar with. They also both challenged the design and layout of the CCO diagram, which has essentially remained unchanged since its origin in 1998.

A comparison between feedback from the two evaluations with the browser-based prototype can be made. Whereas naturally usability experts provided very specific ideas about improving the general usability of the prototype, feedback from SMEs was much more focused on the actual way CCO is used. As a result of the evaluations, on a number of occasions development priorities were reshuffled.

#	Version	Participants	Group Size	Participant Expertise
1	1st paper	4	4	SME
2	2nd paper	1	1	SME
3	2nd paper	7	2-3	Students
4	5th browser	3	1	Usability
5	7th browser	2	1	SME

Table 14: Summary of conducted formative evaluations. As a result of rapid prototyping, work-in-progress versions of the toolkit were released every fortnight. The 5th such browser-based version was considered to have sufficient functionalities to start formative evaluations.

Some of the ideas that SME participants proposed are commonly used tactics like police patrols or installing CCTV, others are more specifically tailored and sometimes innovative. In this initial set some ideas often get repeated by different players. Of the six instances (four with the second paper prototype and two with the browser-based prototype) when users generated their ideas themselves, there were six ideas that came up three times or more. The pattern matching mechanism managed to match very few of those. As a result scores were not representative in the eyes of participants. However, when one of the participants saw someone else's ideas being the same as the one she had just contributed, she got noticeably intrigued.

6.4.1 Web-based Toolkit

This subsection describes the final prototype version of the toolkit that was used in the evaluation studies. Here, scoring and discussion are built-in, addressing respectively *achievement* and *socialisation* in Yee's motivational factors (2005; 2012). In the process of doing that elements from Fogg's *captology*, or *persuasive technology*, (2009, 2002) can be identified.

The toolkit guided participants through a process consisting of four consecutive stages: introduction to the scenario, idea generation, idea assessment and score review. This sequential process was the instantiation of Fogg's tunnelling. The toolkit is an instance of using *captology* as a *medium* in that it helps learners understand cause-and-effect relationships. This is illustrated in detail in the following evaluation section.

The scenario part essentially provides learners with the opportunity to again read the scenario that they already encountered in the introduction to the evaluation sessions. At the end of this step, as in general with the process, they were able to determine for themselves when they were ready to move on to the idea generation part. This was meant as a form of personalisation by making the toolkit adapt to the learner's own pace.

After learners are introduced to the scenario, they are taken to the subsequent idea generation part. It features an interactive version of the CCO diagram. It is provides in two consecutive modes – identification of *causes* and suggestion of *interventions* (the latter being shown in Figure 15). Each of those allows learners to focus on one single ray of the diagram by clicking on it. When a learner clicks on one of the 11 contributing rays of CCO on the causes diagram, a dialogue box appears, that provides an explanation and an example for that particular causal ray. In this dialogue box learners are provided with the opportunity to write down suggestions of how this generic cause is being instantiated in the scenario at hand. The emphasis on cause or intervention rays in the diagram to allow learners to focus on one ray and provide ideas only for it are instances of how the toolkit simplified participants' work with the CCO framework.

Similarly, in the interventions diagram a dialogue box provides context to learners by reminding them of their suggested causes, and typical intervention principles relevant to the corresponding generic cause. Typically intervention principles belong to one of the 11 CCO causal rays. Learners suggest their own methods and match them to the

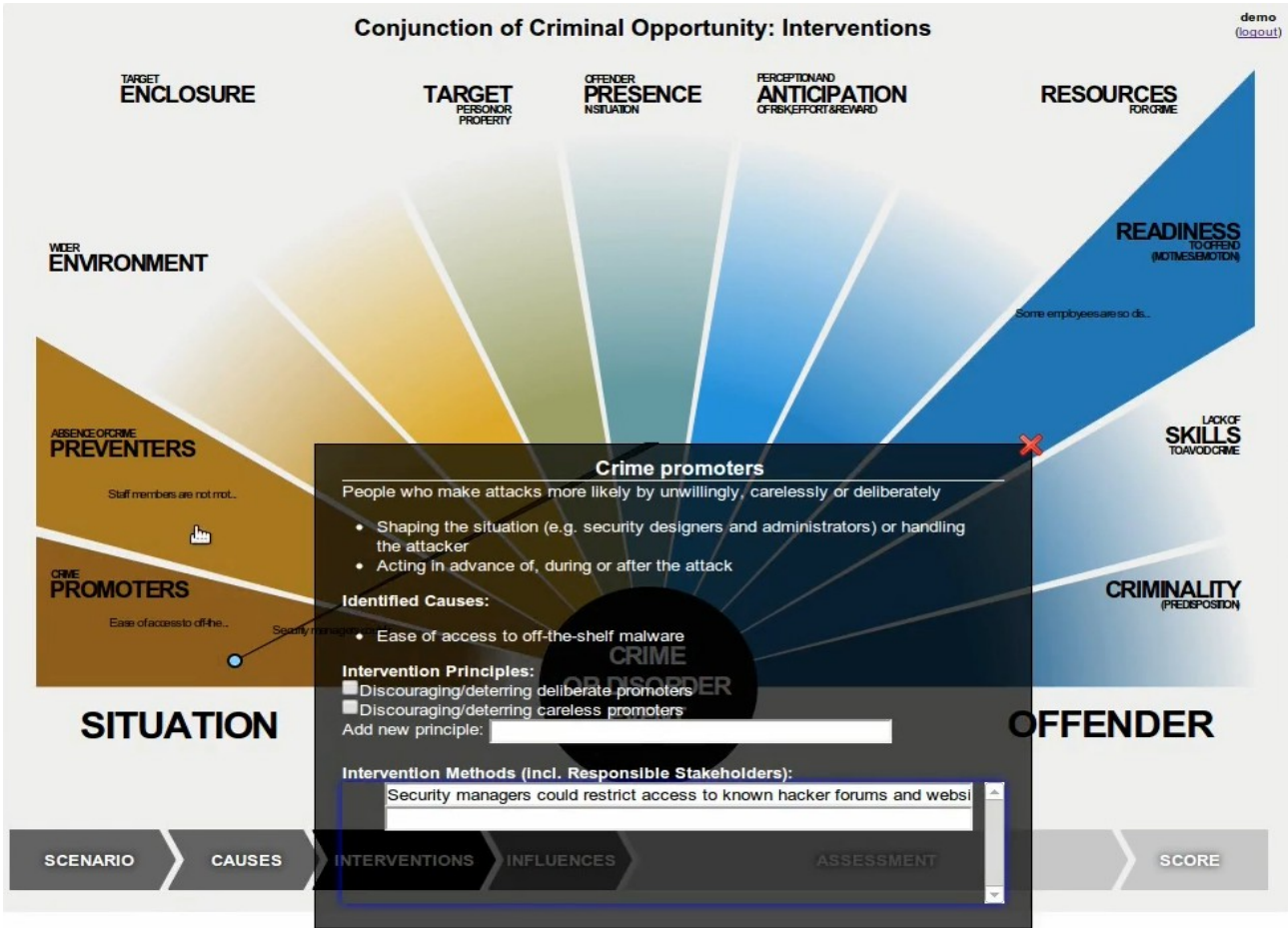


Figure 15: The idea generation screen, as recreated by the toolkit for the identification of interventions. It features visual feedback for identified ideas, and semi-structured input dialogue for participants to suggest ideas.

corresponding list of principles. Learners also have the opportunity to add custom principles to any of these 11 lists. In the background the toolkit employs a simple word-based pattern-matching algorithm (Weiss et al 2004) to try to match newly suggested ideas to existing ones. When the algorithm cannot match a new contribution to any existing with sufficient certainty, the new idea is annotated as innovative and a "new" icon (also known as toast, because of the visual resemblance to toasting slices of bread) pops up on the learner's screen. This icon is intended as a form of immediate praise to participants that come up with new ideas.

After these phases of generation learners are given access to a table with the proposed interventions as rows and the 11 causes as columns (see Figure 16). This gives them the opportunity to identify possible matches between any of their suggested interventions and the generic causes. This way they can further explore the influence a suggested intervention could have on wider causes and correspondingly how it is interconnected with other interventions.

Both the set of causes and interventions diagrams on one hand, and the influences table are examples of exploring causes and effects using *persuasive technology* as a *medium*.

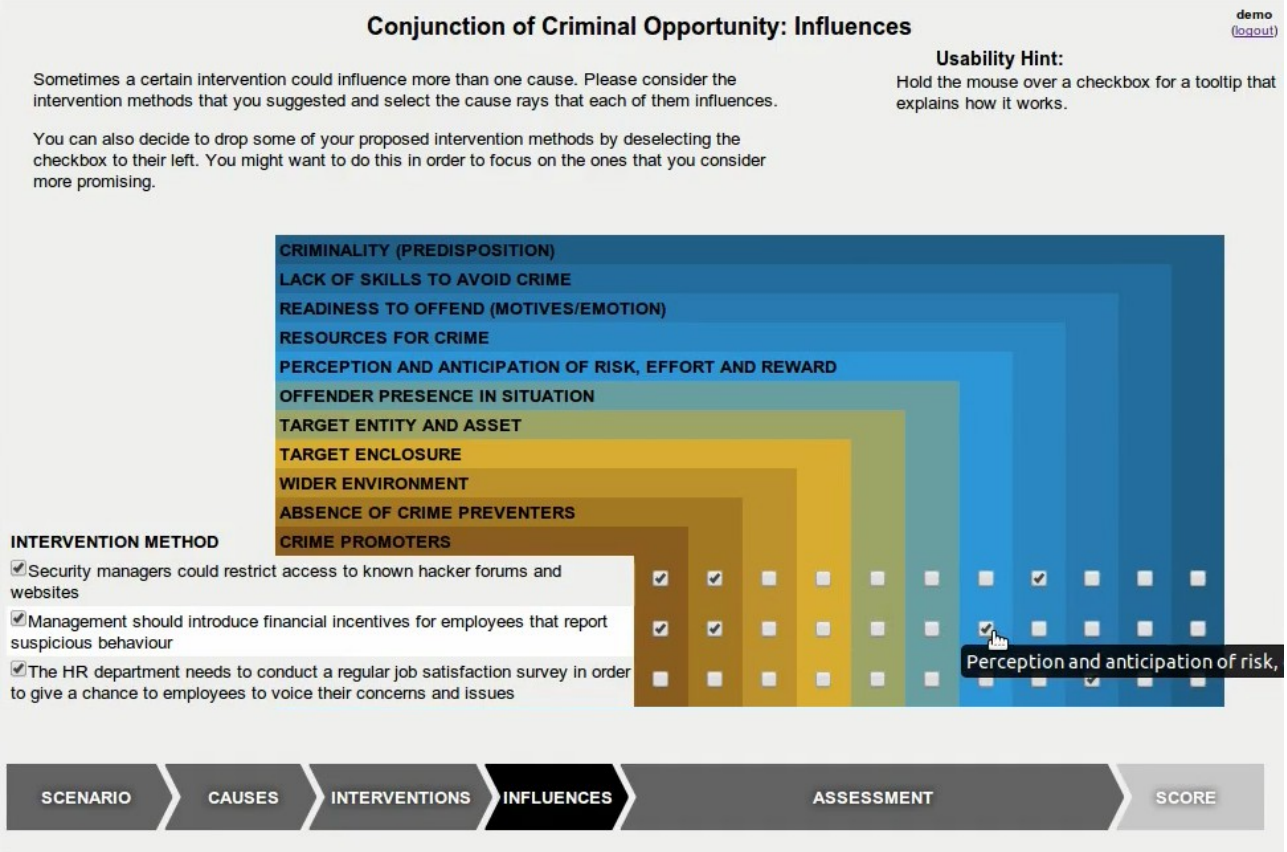


Figure 16: The toolkit screen that asks participants to think whether their proposed interventions could influence other causes beyond the one that they were originally designed for.

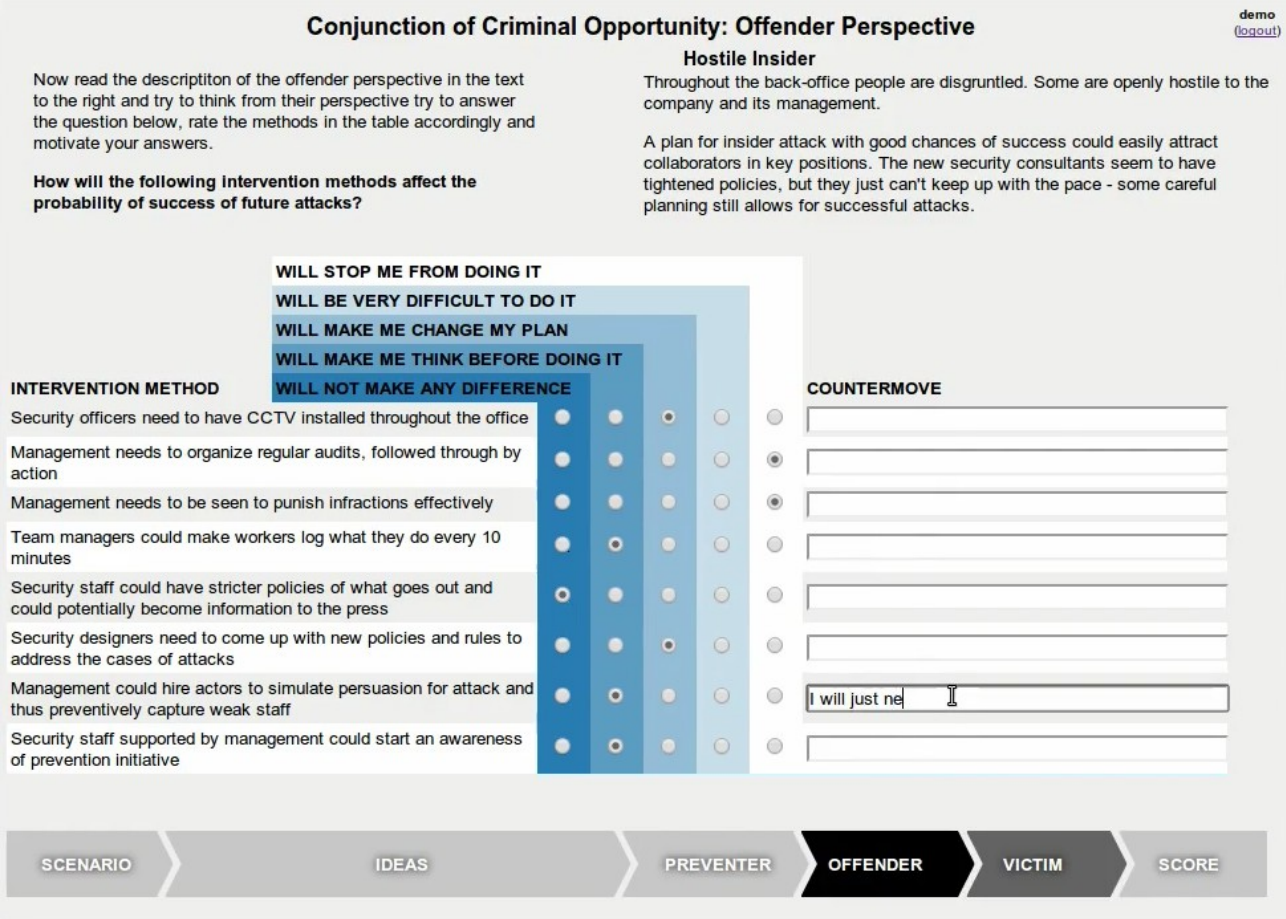


Figure 17: The second of the three assessment screens. Here participants are asked to assess how interventions will affect expectations of success by offenders.

After identifying further influences the toolkit engages learners in a role-based assessment of the proposed interventions. Here, for the purposes of exploring variation, the toolkit provides two alternative workflows that can be delivered to study participants using it. The two alternatives deliver to learners different sets of ideas to review in their assessments. With the control group workflow leads participants to assess their own method ideas, whereas the one for the experimental group asks for assessment of a predetermined set of ideas, meant to be seen as someone else's contribution. Learners are asked to grade each idea along a 5-point Likert-scale and are provided with an empty text field if they wanted to provide further comments or questions about that particular assessment. This is where evidence of shaping of learning goals (as explained by (Bransford and Schwartz, 1999) and reflective discussion occur.

Each learner makes an assessment from three different perspectives. In the first and second screens of the role-play learners rate how the implementation of each intervention idea would affect the probability of further attacks. Learners are first asked to assess how strongly each idea impacted the chances of the criminal opportunity. After this initial

assessment, two more follow, this time asking learners to think from a different role perspective. One of the subsequent assessments puts them in the role of the offender (Figure 17) assessing how ideas impact their perception of the criminal opportunity. In the last assessment learners are put in the role of neutral citizen who could potentially become victim of crime, helpful preventer or even unintentional crime promoter. In contrast to the previous assessments where probability of criminal events is explored, in the third one learners are asked to assess how far intervention ideas may have reduced the harm that the criminal event might cause. These assessments, as well as the score and ranking screen that is to be explained subsequently, enable the toolkit to provide learners with a sense that the technology suggests *reciprocity*.

In the background the toolkit uses a simple pattern matching mechanism to cluster intervention ideas. It is data-driven and ideas are represented in a *bag-of-words* form, only counting word occurrences. This representation is used to define aggregated word clusters and subsequently to match new ideas towards the average word usage. This entire process makes it possible for the toolkit to provide feedback on ideas, suggested by learners, and to ultimately assign a score to learners using the toolkit. The idea matching mechanism also allows for better chances of identification of innovative ideas – those that do not match any existing pattern.

All this is fed back to users in the final score and ranking screen (Figure 18) that shows to learners their performance. This includes a table with intervention ideas, suggested by the learner, and overall statistics and ranking of their performance. The table features a breakdown of the three scores these ideas have cumulatively received from other learners and feedback from previous learners provided via comments they gave while assessing previous similar ideas.

This last screen also summarises these scores into learners' overall score and ranking. Learners see their provisional ranking of their corresponding experimental group according to three distinct metrics: i) overall score; ii) average score per intervention method; and iii) number of proposed innovative method ideas (the ones that the pattern-matching algorithm could not identify with existing ones).

Delivering assessment and then seeing how this was reflected in the final score was also an example of using persuasive technologies as a medium.

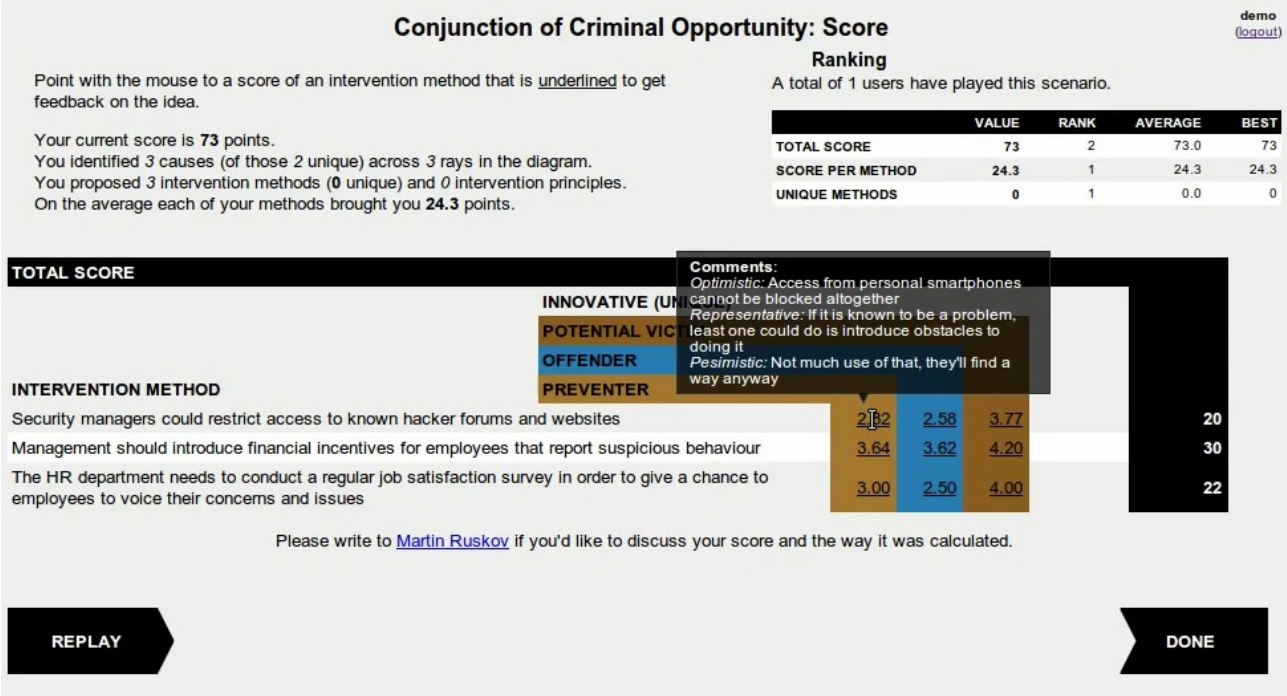


Figure 18: The final score screen. In the upper part of the screen are scores and rankings. When a participant hovers over an individual assessment of an idea with the mouse, a pop-up with the comments given to this assessment appears.

6.5 Conclusion

The adoption of a rapid prototyping approach gave early feedback and guidance in the development process. Resulting from this approach, evaluation activities had to be adapted accordingly. Taking a more flexible approach towards evaluation allows for earlier insights into the implementation of certain design ideas and consequential early remedial actions in the game design.

This development focused on delivering a software toolkit satisfying only a minimal set of requirements that would enable easy learning of a complex crime prevention framework. Early feedback that I received indicates that utilising interactive diagrams where players can drill-down into complex aspects is a promising way to do that. I've also collected indications that gamifying the process further engages some of the users.

The idea that aligning game topic and in-game examples focuses contributions towards the intended knowledge-sharing objectives was reconfirmed by feedback from users of the CCO prototype. They were very specific in providing feedback on the visual representation of both case and theory, as well as the terminology being used. Those users wanted their environment to be visually attractive and believable.

Furthermore the multitude of commonly recurring ideas suggests clustering of player ideas might be relatively accurate. This is possible, because the prototype puts contributions in

the context of CCO and its scenario. As a result, wordings are easier to interpret contextually.

Despite its existence in over a decade and positive reception among crime prevention practitioners across the world, feedback collected in these studies, challenged CCO itself and its application to information security. Participants had very specific comments regarding the visual representation of the theory and the terminology used. As a result, efforts have started towards re-phrasing and re-designing the CCO diagram and its terms, but this theoretic and knowledge-management process might continue beyond the scope of the game prototype project.

Chapter 7: Lab Evaluation of the CCO toolkit

Once a satisfactory prototype was developed the research continued with a study exploring the learning effects when using it. In this section I report a learning study conducted with the toolkit described in the previous chapter. The aim of the toolkit is to improve awareness of information security problems and solutions. In this particular study the objective was to explore how the toolkit helped non-experts improve their understanding of information security (RQ1), improve their ability to apply what was learned (RQ2), and finally, improve their engagement in the learning process (RQ3). This was to be achieved by using variation to make them think in depth of the risks that their behaviour might expose their organisation to. The scenario used describes people working with valuable business information employed by organisations. The intention is to engage learners in working towards being active preventers of information security threats and not inadvertent promoters.

As the autonomous use of serious games has already been acknowledged as a process of interest, I wanted to encourage participants to also use the toolkit remotely. This would allow for a study of self-paced learning with the toolkit. Due to the complex nature of the toolkit, this would ideally happen after participants were already familiar with it. That's why participants were invited to take part in a remote access study, in the form of a competition after the lab sessions.

Here I report a lab study that I conducted to evaluate the toolkit and explore learning variation.

7.1 Materials

The materials that were necessary for this study were all embedded in a bespoke website, featuring the *CCO toolkit*. These were an introduction to the CCO framework, the tailor-made problem scenario and the actual web-based toolkit used to navigate participants through the framework, to provide the necessary guidance and feedback and to actually collect the user-generated data. There was also the learning assessment test deployed in a popular online survey system – also integrated with the website.

7.1.1 CCO Introductory Text

CCO is a multifaceted and complex framework. For the purposes of the study I had to compose a written introduction that would be short enough to fit on one page – to ensure

it did not take up too much time within the study sessions or deter participants. Faced with this challenge, I decided to focus on information that would not be easily interpretable from the toolkit itself, while at the same time providing the necessary background knowledge and awareness to allow for a quick start (see in Appendix C).

The eleven causal elements of CCO can be read off the diagram when participants use it. The distinction between principles and methods, on the other hand, is better understood when grounded in examples. That's why a decision was taken for a functional description, rather than a factual one. The description was based on the preventive process (Ekblom, 2011a), as it gave an overarching view of how causes and interventions (the objects of interaction) fit in the bigger picture. This description did not include the eleven elements, nor did it include the diagram, details about the use of the diagram, or explanation of intervention principles and methods. As described previously in this chapter, these were left to the toolkit to take care of.

7.1.2 Problem Scenario

In order to give participants the opportunity to practise application of the framework, the scenario had to introduce them to a realistic description of a recurring problem. This had also to be a situation without a seemingly straightforward solution that would have left them with the feeling that they had solved it in just a few minutes.

The scenario that I developed was based on survey data from the CERT Guide to Insider Threats (Cappelli et al., 2012). In order to make it representative, I designed it to apply to two of the three most commonly recurring sources of insider threat: IT sabotage and theft of intellectual property.

The scenario describes a frequently relocated outsourcing centre for IT services where disgruntlement among staff grows to the extent that rapid turnover leads to a hit-and-run culture of insiders attempting to make a big win at the company's expense. The text featured examples of plausible cases of insider attacks within the scenario, intended to illustrate their diversity. These ranged from numerous activations of virus protection software to leaking sensitive data or poaching customers when leaving the company. The full text is available in Appendix C.

7.1.3 CCO Toolkit

Beyond the introductory text and the problem scenario, the toolkit was used as described in Section 6.3.1. For this study the separation into control and experimental groups was

employed. Once again, the control group got to rank and comment on their own ideas for intervention methods in the assessment phase, whereas the experimental group got a predefined set of ideas to assess. One consequence of this differentiation is that was the difference in number of ideas participants in the two groups had to assess. The experimental group received 8 ideas in an assessment perspective, thus a total of 24. Each participant in the control group, in contrast, had to assess as many ideas as they suggested, and had to assess the same ideas from the three different perspectives. Thus the number of ideas that participants in the control group had to assess varied according to their activity in the lab session.

7.1.4 Assessment Materials

A portfolio of learning assessment measures was designed in order to capture progress

Code	Question	Addressed ILO
TQ1	How would you describe the CCO framework?	ILO1
TQ2	What is the CCO framework used for?	ILO1
PQ1	What are the key causes of the insider attacks in the above scenario?	ILO2
PQ2	What are possible interventions methods that would reduce or prevent further attacks of this sort?	ILO4
PQ3	For each of the methods, please suggest one reusable principle that generalises the approach that has been used.	ILO3
MCQ1	Which of the following (if any) are causes working on an employee at a bank to help the commitment of financial fraud?	ILO5
MCQ2	Which of the following actors (if any) have interest in secretly planting a trojan onto a home computer?	ILO5
MCQ3	Which of the kinds of methods below apply to a “use secure password on your private computer” publicity campaign within a company?	ILO3
MCQ4	Which of the following (if any) could be parts of the enclosure around a file that is potential target?	ILO1, ILO5
MCQ5	At an open access internet café which of the following (if any) are potential non-professional crime preventers?	ILO5
MCQ6	Which of the following (if any) are well-formulated intervention principles?	ILO5
MCQ7	Which of the following (if any) are intervention methods rather than intervention principles?	ILO5
MCQ8	Which of the following (if any) are resources for a potential offender to commit an insurance fraud?	ILO5
MCQ9	An IT company has several cases of intellectual property leaks to competitors. For which of the following (if any) could they use the CCO framework?	ILO1

Table 15: Questions used in the assessment of this study and corresponding intended learning outcomes.

	Question	Code	Construct
iGEQ1	I was interested in the game's story	IMM	Sensory and Imaginative Immersion
iGEQ2	I felt successful	COMP	Competence
iGEQ3	I felt bored	NAFF	Negative Affect
iGEQ4	I found it impressive	IMM	Sensory and Imaginative Immersion
iGEQ5	I forgot everything around me	FLOW	Flow
iGEQ6	I felt frustrated	TEN	Tension
iGEQ7	I found it tiresome	NAFF	Negative Affect
iGEQ8	I felt irritable	TEN	Tension
iGEQ9	I felt skilful	COMP	Competence
iGEQ10	I felt completely absorbed	FLOW	Flow
iGEQ11	I felt content	PAFF	Positive Affect
iGEQ12	I felt challenged	CHAL	Challenge
iGEQ13	I felt stimulated	CHAL	Challenge
iGEQ14	I felt good	PAFF	Positive Affect

Table 16: Items in the iGEQ standard measure.

corresponding to each of the intended learning outcomes, and respectively to Sugrue's knowledge taxonomy. The measures included two theoretical open-answer questions to explore participants' (learners) understanding (see TQ1 and TQ2 in Table 15, addressing RQ1), three problem-specific open-answer questions to test their ability to apply the framework (PQ1, PQ2, PQ3, see Table 15, addressing RQ2), and nine multiple-choice questions to test their general understanding and their ability to transfer what was learned to other contexts (MCQ1-9, which is also a matter of application, or RQ2). Each of the multiple choice questions included four possible answer options and participants were allowed to select any number of correct answers, or none. This was different from the *vLeader* study where Likert-scale questions were used. However, it was considered that multiple-answer questions are more suitable for learning assessment, as their answers are more objectively determinable, when compared to a Likert-scale. Participants were also given the opportunity to provide further comments or clarifications regarding each possible answer.

Participants were asked to fill the 14-item in-Game Engagement Questionnaire (Ijsselsteijn et al., in press; van den Hoogen et al., 2008) after use of the toolkit as a measure of their engagement (RQ3). They had to answer the iGEQ (as it is abbreviated) questions using a 5-point Likert scale ranging from "Not at all", "Slightly", "Moderately", "Fairly" and

"Extremely". The questionnaire plots participant responses along 7 dimensions, each addressed by two questions (see Table 16).

7.2 Study Method

To introduce the study, this section reports how participants were recruited, what procedure was employed and how data was analysed.

7.2.1 Participants

For this study 28 participants were recruited from a university recruitment pool and offered a financial rewards, that had a base component and grew if they were in the top three performers in any of the three game score metrics. Of them 19 were male and 9 were female. Their age ranged from 20 to 65 with an average of 26.5 and median 23.5. Five of those participants reported that they had some previous exposure to information security or a related field. In general this was only limited and ranged from training on information security in the army or at university, to deploying firewalls and anti-virus software. None of them had actual professional experience.

7.2.2 Procedure

For the purposes of the experiment participants were split into control and experimental conditions, 14 participants each. When invited to the experiment by e-mail, participants were given two one-page-long texts to digest. One of these was an introduction to the CCO framework and the other was a problem scenario, describing an insider threat and set out in detail below. Both of these texts were available to participants for reference throughout the study. A total of three tests were conducted – two in the lab session with one before and one after use of the toolkit; the third test was administered remotely six weeks after lab sessions.

The laboratory session lasted 90 minutes and involved using the toolkit for up to 60 minutes with two tests. Time was planned so that they finished at least 10 minutes before the full 90 minutes were over, to give them time to do the final test. After that test participants were interviewed about their experience. They were also invited to take part in a subsequent competition, in which they had the chance to use the toolkit further and the ones that ranked first were offered a cash prize. At the delivery of the prize further interviews with the best performers were conducted. Details of this procedure are explained below in their order of occurrence.

The tests (described in Section 7.1.4 above) aimed to assess participants' knowledge of crime prevention and the CCO framework in particular (RQ1), in the course of analysing the causes of the exemplar crime problem and coming up with, then assessing, preventive solutions (RQ2).

The experimental part of the study was used to examine the effects of variation on learning with the toolkit. As part of the process guided by the toolkit the experimental group was asked to assess a predetermined set of ideas, presented as if other study participants had written them, whereas the control group merely assessed their own ideas.

When using the toolkit participants were deliberately not given any additional instructions to the ones present in the toolkit, unless they asked for them. If they did not suggest any ideas, they were told they needed to come up with at least two to continue. In general they were not reminded of time, unless they took too long in the first idea generation steps (refer to the explanation of the toolkit for the full process in the materials section below). If this happened, I, as the researcher present in the room, told them that there was more to do in the toolkit and that they should finish with the idea they were currently writing down and proceed with subsequent steps.

The test after using the toolkit also included iGEQ – a 14-item Likert-scale questionnaire measuring participants' perceptions of engagement (Ijsselsteijn et al., in press; van den Hoogen et al., 2008). This was one of the two measures intended to address RQ3.

In order to encourage participants to use the toolkit from home a competition and prize rewards were announced. The competition utilised the three different score rankings implemented in the toolkit. As explained in Section 6.3.1 these were best total overall score; best average score pre method; and most innovative (as recognised by the toolkit) ideas. As a way to allow broader competition, the top performers on each of these three different rankings were rewarded. The number of uses of the toolkit after the lab session was used to measure the continued engagement of the toolkit (RQ3).

To encourage participation in the retention test, completing the test was a prerequisite for a participant to take part in the competition. Due to low participation, with the last reminders for the survey, further rewards for second and third place on each of the three toolkit scoring metrics (described at the end of Section 6.4.1) were announced.

The qualitative part includes analysis of the two interview sessions – at the lab session and the subsequent later interview. Both interview sessions were recorded and the researcher

1	How did you find the prototype? Did you feel comfortable in using it?
2	Do you feel that you know what the Conjunction of Criminal Opportunity? Do you feel confident that you know how and when to use it?
3	How do you think your awareness of information security changed?
4	What was more attractive and what was less attractive about the prototype? Is there anything particular about the task that interested you?
5	Was there anything that you found particularly confusing? Was there anything particularly difficult?
6	What do you think about the methods and comments that you were given in the prototype? Were they useful? Did they influence your thinking? Did they help you improve your own ideas for methods?
7	In one sentence, please describe what you learned from the prototype.

Table 17: Questions asked during the in-depth interviews at the end of study Study 3. These questions are indicative and were adapted for each individual student.

took notes. Similar questions were asked in both sessions, based on the prototypical questions listed in Table 17. The actual questions were adapted, based not only on interview session, but also on changes considered the level of engagement and their experience according to experimental condition, among other factors.

7.2.3 Analysis

As already mentioned the wider study was a controlled experiment. The difference between control and experimental conditions was that participants in the former were required to assess the ideas they had previously generated *themselves* on causes and intervention, whereas those in the latter were shown ideas of *someone else* to assess. This is further clarified in the section dedicated to the toolkit in the materials section below. The distinction between the two conditions is of minimal relevance for this piece of analysis – for present purposes the results reported for both conditions were equally valid and hence combined here.

The application questions (PQ1-3) were analysed in two different ways – quantitatively and qualitatively. The quantitative evaluation covered only the suggested intervention methods (PQ2). Independent experts (postgraduate students in Information Security) were asked to rate all ideas for methods suggested during the tests. To do this, methods from the entry, exit and retention tests were randomly shuffled and fed into a modification of the assessment part of the toolkit. In contrast to the way study participants made assessments (see toolkit description in Section 6.3.1), experts made only one assessment per idea, thus this had to be holistic. Experts were asked to assess the impact the intervention might have on the problem along a 5-point Likert scale. The scale featured the ratings “1. Has no

impact on problem", "2. Slightly reduces problem", "3. Partially reduces problem", "4. Significantly reduces problem", "5. Completely resolves the problem" and a free-entry text box for comments. The fact that ranking data is *ordinal*, led to the use of Kendall's W with correction for ties (Kendall and Gibbons, 1990) as inter-rater reliability measure.

Questions PQ1-3 were also analysed qualitatively by means of thematic analysis. In the responses to these questions cause and intervention ideas are problem-specific and thus difficult to assess objectively even by a domain expert. This is further complicated by the fact that participant responses are short and sometimes feature certain ambiguity. As a way to overcome this, the suggested ideas in answers were coded (and thus generalised) into derived broad categories. PQ2 and PQ3 addressed two aspects of the same issue, respectively the methods and the principles of a small set of interventions, so they were analysed together.

The theoretical questions were analysed both with thematic analysis and with the SOLO taxonomy, the former used to explore the particular themes, and the latter – their interrelatedness. The transferability questions were statistically analysed using unpaired samples t-test with the assumption of equal variances.

Thematic analysis was also used to code the interview data. This was done based on the researcher's notes. The codes were subsequently expanded with partial transcriptions from the interview recordings. Finally, following Laurillard's conversational framework (2001), the interviews were interpreted to compare participants' understanding against the original CCO framework.

When reporting quotes from the interviews codes were used to anonymise participants. The codes would typically consist of a letter indicating the study group (E for experimental and C for control), a double-digit participant number and a letter indicating when the quote was made (I for immediate, and D for delayed). A typical example for a code would be C04D, indicating that the participant assessed their own ideas with the toolkit, and that the quote is from their second interview.

7.3 Results

There are three types of study results reported here: quantitative summary of participants' engagement with the toolkit, learning assessment of responses to the corresponding materials and quotes from the interviews conducted after the study intended to elicit participants' perceptions of the experience. The first part shows the extent to which

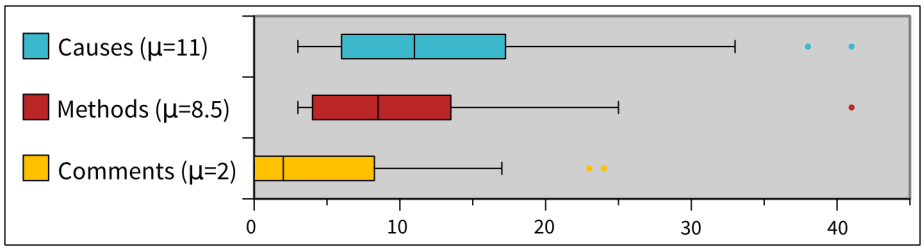


Figure 19: A box plot showing the number of user contributions when working through the toolkit in the lab session. μ indicates the median of each value.

participants engaged with the toolkit; the second, the strength of evidence for learning; and the third, their perception of the toolkit and the CCO framework.

7.3.1 Engagement

This subsection reports results from the analysis of user ideas, generated in the toolkit and from the iGEQ questionnaire.

User-generated Data

All 28 participants in this study were able to effectively identify causes and propose interventions, even though they were not security experts. Figure 19 and Figure 20 show the number of ideas and comments that participants generated during their use of the system. On average participants came up with 11 causes, 8 intervention methods and two comments. Typically participants wrote one cause per CCO causal ray on the diagram. Some participants filled all comment fields, but 11 participants did not fill any. Two-thirds of participants generated more than 20 ideas and comments each.

Typically the variation group came up with more comments (on average 7.5 per person) than those that commented on their own ideas – the control group (average: 3.2 per

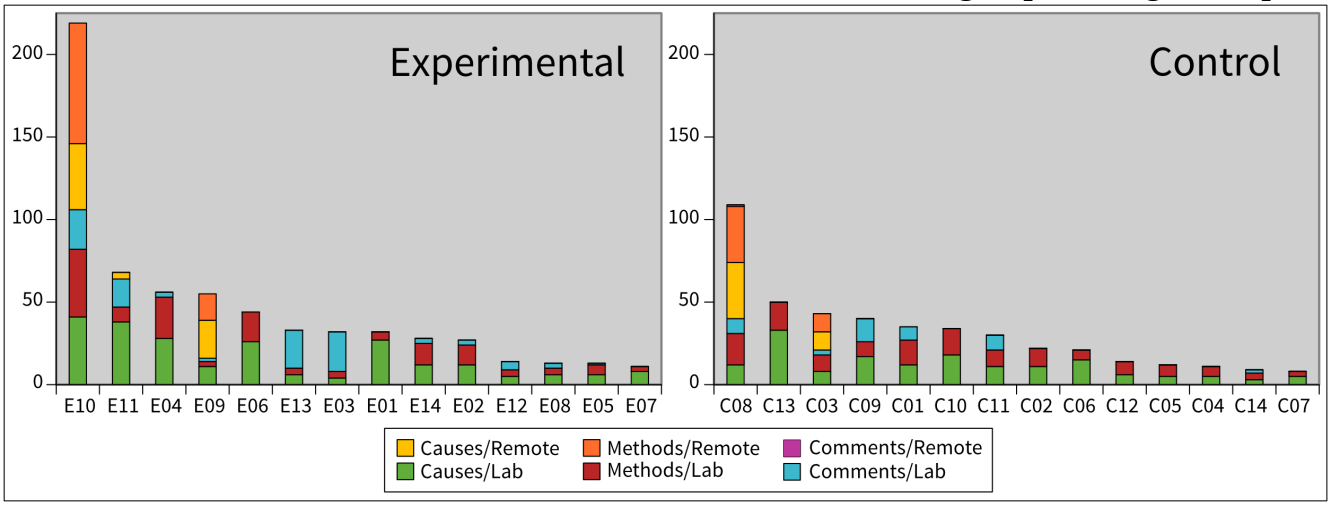


Figure 20: Engagement of individual study participants. On the left is the variation group and on the right is the control group. The different types of ideas are separated according to when they were generate – during the lab session or during the subsequent remote use.

Code	Construct	t(df=26)	p-value	Control Mean	Experimental Mean
IMM	Sensory and Imaginative Immersion	0.88	0.1946	6.79	6.36
COMP	Competence	0.26	0.6	5.64	5.5
NAFF	Negative Affect	-0.88	0.8068	4.14	4.71
FLOW	Flow	0	1	6	6
TEN	Tension	-1.16	0.1277	3.36	4.29
PAFF	Positive Affect	0.84	0.2041	6.43	5.93
CHAL	Challenge	1.67	0.0531	7.86	7

Table 18: Statistical analysis of engagement questionnaire.

person). Participants in the variation group had the opportunity to comment on ideas of others, and 4 participants attempted to provide comments on all 24 provided ideas. 2 of them carried this through the remaining assessment screens. Only one person in the control group made a corresponding attempt, but only completed the first half of the provided comment fields. While in the control group more than 50% (8 participants) did not find it necessary to provide any comments, in the variation group all but 3 participants left at least one.

There was only partial involvement in the remote-use part of the study. Only 12 participants filled the retention test and only 5 contributed (C03, C08, E09, E10, E11) with new ideas after leaving the lab. These numbers did not allow for statistical analysis to draw conclusions about RQ3, but participants were questioned about the reasons why they played remotely or not. Responses are summarised in the consequent subsection dedicated to the interviews.

iGEQ

In Table 18 are constructs derived from the iGEQ questionnaire and two-sample t-test significance levels, assuming normal distribution of data (df=26).

There were no statistically significant differences to support RQ3, even when a one-sided t-test with mean assumption derived from the sample was run. Nonetheless, worthy of notice is a result that is close to the $\alpha = 0.05$ confidence threshold. The control group seemed to be more likely to name their experience *challenging*.

7.3.2 Learning Assessment

Results from the learning assessment are reported here (due to the fact that only 12 of 28 participants took the retention test, it is not included). This section addresses qualitative

Id	Before		After	
	Idea	**Class**	**Idea**	**Class**
C01	*Keeping silent over number of attacks.*	interpret	*Relocation to tax haven where policing is less.*	interpret
	Poor upkeep of software quality.	own	*Staff complaints over pay structure.*	interpret
	Concerns over pay.	rephrase	*Management's poor, uncaring image.*	own
	Rapid turnover of staff.	rephrase	*Poor control over sensitive data.*	own
	Relocation to tax havens.	scenario	*Poor procedure for backups when data servers are down.*	own
V07	*blackmail received about software*	examples	*blackmail*	examples
	house data servers being down	examples	*Bad servers*	own
	shelf malware.	examples	*not activating your antivirus programme*	own

Table 19: Ideas for causes (answers to PQ1) suggested by two participants as an illustration of the classification of ideas. All of these participants are provided.

and quantitative analysis of the practical questions (providing insights into RQ2), qualitative analysis of the theoretical questions (RQ1) and quantitative analysis of the multiple-choice questions (RQ2).

Practical Questions

The generalised categories for causes identified by participants (PQ1) can be considered to represent some sort of continuum depending on how close to the original scenario text they were. The considered categories are exemplified in Table 19 and explained below:

- examples – causes that directly reflected examples provided in the scenario (see Appendix C for the exact description)
- scenario – causes that were suggested by the scenario (or the subsequent role descriptions)
- interpret – causes that are not based on the scenario, but could be related to something explained in it
- rephrase – causes that are based on the scenario, but a different wording is used, and sometimes this results in some change of meaning
- own – causes that are not directly explainable by the scenario and are proposed by participants as their own ideas

Figure 22 shows the proportion of ideas before and after exposure to the toolkit. From that

figure it is evident that there was a shift from causes implied by the scenario or variations of them to more indirect and complex causes that were original suggestions of participants. The average proportion of own ideas almost doubled from 19% to 36%.

When considering interventions (methods and principles grouped together), the groups listed below were proposed and are exemplified in Table 20.

- leadership – interventions that addressed the need for better leadership as a way to consolidate the team
- training – interventions that had to do with providing employees with opportunities to improve
- paystructure – interventions that had to do with the improvement of the pay structure for employees
- whistleblowers – interventions that had to do with encouraging employees to report suspicious activity around them
- punishment – interventions that had to do with penalising attackers or promoters
- narrative – interventions that had to do with the company trying to influence employee perceptions
- office – interventions that had to do with changing the physical space or its location, excluding immediate security measures
- security – interventions that were a direct form of security improvement

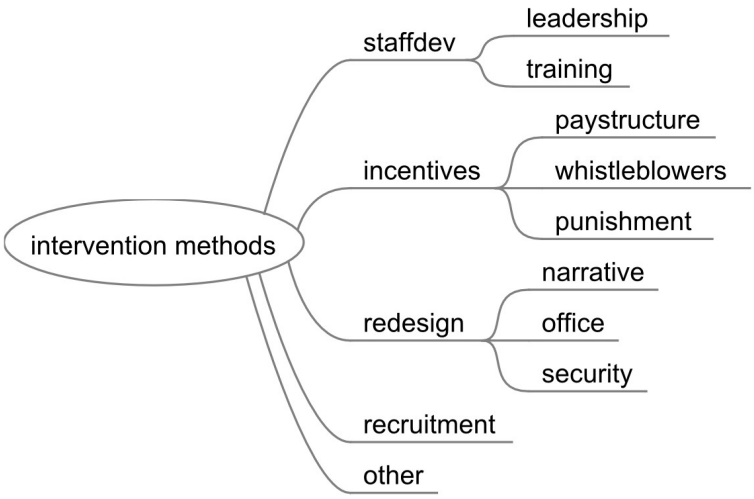


Figure 21: The hierarchy of ideas, as initially classified and subsequently grouped to allow for better oversight

- recruitment – interventions that affect the employee recruitment process
- other – other unrelated ideas altogether representing less than 3% of what was proposed

A further level of abstraction allowed ideas to be grouped as illustrated in Figure 21. The resulting final groups are the ones used to represent comparison of outcomes in Figure 22.

The proportion of interventions (methods – PQ2 and principles – PQ3) that have to do with staff development increased from 22% to 28%. Such method ideas had to do with solutions such as training or showing better leadership, which are essential to security issues related to disgruntlement (Kirlappos et al., 2013). The increase affected ideas related to both leadership and training, and was notably at the expense of ideas that fell under the classes of company-driven narrative of the situation, punishment and recruitment.

In the expert assessment experts provided assessment for 240 ideas. These were 98 ideas from each of the entry and exit tests, and 44 from the retention test. Agreement between the two involved experts was low (see Table 20), so I decided to involve a third expert. This did not result in better agreement between any two of the three experts, but comparison of all three ensured greater confidence.

Independently from the lack of agreement between experts, their individual assessments of participant intervention methods did not provide any indications of improvement immediately after using the toolkit. Table 21 shows averages and standard deviations of the scores participants achieved, according to the three experts. It also shows the high percentage of ideas that experts felt needed clarification beyond the scalar assessment.

Theoretical Questions

In responses to TQ1 several participants provided reasoning for the choice of 11 generic

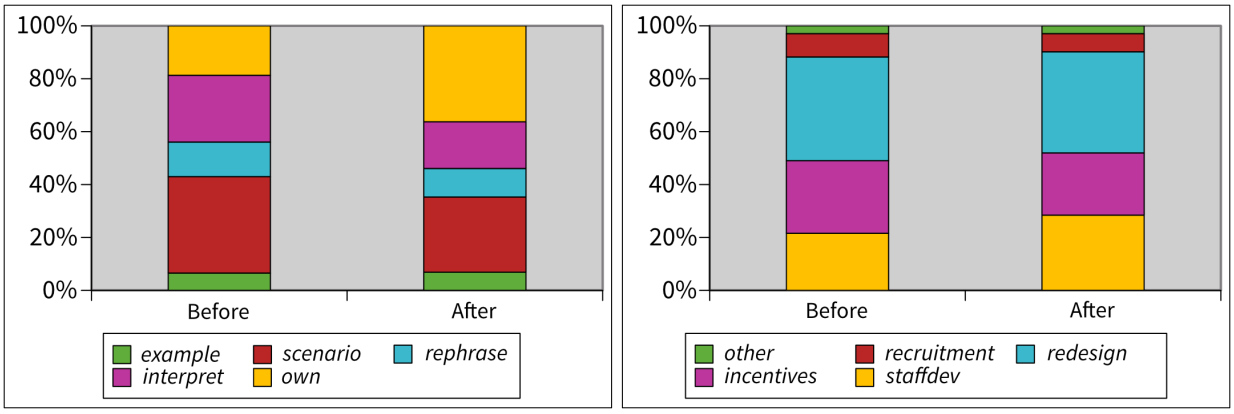


Figure 22: Distribution of answers to application questions: left - PQ1 and right - PQ2 and PQ3 (same distribution). The graphs compare answers before and after engaging with the game-based toolkit.

Raters	Wt	Chisq(239)	p-value
SE1 and SE2	0.605	289	0.0143
SE2 and SE3	0.612	292	0.0105
SE1 and SE3	0.585	280	0.0361
SE1, SE2 and SE3	0.468	335	0.00004

Table 20: Statistical analysis of inter-rater reliability of experts assessing intervention ideas.

causes. One example illustrates how closely that reasoning follows that of the framework originator's: "*The CCO framework focuses on 11 key, direct causes of crimes, that apply to victims, offenders, and preventers. All other, "higher" causes, such as socioeconomic factors, can be though of as acting through one or more of these key causes.*" (E02) Participants were not informed about this, but this is precisely how CCO is intended to be used and in fact another version of the CCO diagram represents causal distance (which is direct vs remote causes) along the rays (Ekblom, 2011a).

Most participants showed some form of learning in their answers to the theoretical questions (TQ1 and TQ2) when comparing answers before and after using the game-based toolkit. These included rephrasing or relating to previous knowledge, explaining new themes within the subject, integrating different themes learned with the toolkit, or adopting the professional language of the toolkit. However, as can be seen in Table 22 these pieces of evidence were diverse.

Commonly, study participants expanded their answers after using the game-based toolkit, thus showing what new understanding they had developed. Typically in such situations they provided further necessary details in their responses after using the toolkit. In Table 23 E11/TQ1 is an example of a participant not giving anything specific in their answer before using the toolkit, but adding relevant concepts after that. This is an example of someone reaching the unistructural level in the SOLO taxonomy in their answer after using the toolkit. On the other hand C11/TQ1 exemplifies a second aspect (interventions) being added to the one already discussed (causes), thus exemplifying a newly developed

Rater	Before			After		
	Average	Standard Deviation	Comments	Average	Standard Deviation	Comments
SE1	3.40	0.82	27.45%	3.17	1.02	30.39%
SE2	2.92	0.82	45.10%	2.83	0.89	42.16%
SE3	2.53	0.91	40.20%	2.36	0.84	29.41%

Table 21: Average score per intervention method idea in different tests, as assessed by each expert.

Id	Before		After	
	Idea	**Class**	**Idea**	**Class**
C01	*Locate back in original country (or at least not in a poorly-policed tax haven) this is responsibility of management.*	office	*Improve management reputation, talk with employees' representatives this is responsibility of management, Human Resources, employees, time planners.*	leadership
	Develop a good relationship with the press, to avoid snooping journalists. This is responsibility of media team and management.	other	*Move away from tax haven, to better policed and audited country. This is responsibility of management, estates and planning.*	office
	Tighten staff controls, putting staff directly responsible for security on higher wage and longer contract. This is responsibility of Human Resources.	punishment	*Be seen to improve pay structure this is responsibility of management, finance, HR.*	paystructure
	Restrict access to company financial data further. This is responsibility of security and finance team.	security	*Improve deployment of backup servers this is responsibility of IT.*	security
V07	*Always activate your antivirus*	security	*Tighten up the data flow within and outside the company premises this is responsibility of IT, security, management, employees.*	security
	Make sure that your computers firewall is on	security	*Make sure you have a Windows instillation CD, and that you back up all your data*	security
	Don't open emails that you are not aware of that who the sender is.	security	*Update your windows XP software*	security
			updating your antivirus software programme	security

Table 22: Ideas for intervention methods (answers to PQ2) suggested by the same participants as an illustration of the classification of ideas. All method ideas of these participants are included.

multistructural answer.

In other cases the comparison of the two answers showed a change in the way that a given participant considered the topic and demonstrated their ability to better integrate what they had learned. Commonly they first wrote of the process that the toolkit took them through. After using it, they also wrote of its goal or implications, alongside the process (see C01/TQ2 in Table 23 for an example). This is typical for the relational level of the SOLO taxonomy (which was introduced in Section 2.1.3). Among the participants there were no examples of participants moving to an extended abstract level in the SOLO taxonomy.

Participant /Question	Type of Evidence	Before	After
E11/TQ1	SOLO unistructural	*It tries to take a micro-approach in terms of identifying small problems in society that lead to a crime being committed (in terms of Information Security)*	*tries to reduce the risk and occurrence and severity of attacks by interrupting in the causes*
C11/TQ1	SOLO multistructural	*CCO framework is used to reduce crime related to leakage or attack of information by investigating their causes.*	*it is a framework to identify the cause of crime related to information system followed by intervention to eliminate the crime.*
C01/TQ2	SOLO relational	*CCO is used to identify current or potential breaches and to work through all the chains of effect, thus creating watertight solutions.*	*CCO is used to examine the many potential causes of incidents, and to explore what the implications of potential solutions would be, from all angles. Sometimes the implications are massive.*
C03/TQ1	own vocabulary	*a comprehensive method of curbing the occurrence of a particular crime with minimal effect to stakeholders.*	*a comprehensive method of identifying causes, possible solutions and assessing their impact to a particular criminal activity with little impact to stakeholders.*
C13/TQ2	mimicry	*it help people to reduce the chance of being cheated during online security.*	*protect vulnerable people from cyber crimes. give people an insight about increasing cyber crimes.*

Table 23: A selection from the assessment results, illustrating different types of evidence. The third and fourth columns contain unedited participant responses. None of the answers exemplified the extended abstract level of the SOLO taxonomy.

In two cases participants used their own terminology in their answers. For example in C03/TQ1 the participant talks of "stakeholders" and "curbing the occurrence" of crime, but neither of these phrases was used in the toolkit or by the facilitator. This shows that they went through a process of relating what they experienced in the study to what they had previously known, describing the new knowledge in their own vocabulary.

Eight study participants demonstrated no form of learning in either of their answers to the theoretical questions. Instead they provided the same or even less information in their answers after using the toolkit. In two other cases, although participants showed that their understanding had developed in the answer to one of the questions, they only superficially answered the other. In these responses participants used the professional language of the toolkit, but did not provide a response with substantial information in their answers (see C13/TQ2). These cases were considered as cases of mimicry, rather than learning.

Multiple-Choice Questions

Statistical analysis running an unpaired t-test of the transferability multiple-choice

questions (MCQ1-9) showed no significant improvement after using the toolkit (test result: m_{before} = 23.3, m_{after} = 23.9, df=54, t=1.674, p=0.226). Still, on average the number of participant's correct responses improved by slightly more than a half, in other words, every other participant indicated one more correct option after using the prototype. Nine participants provided 21 comments for clarification of their responses to the multiple-choice test before engaging with the toolkit, with one responsible for eight of these comments. Only one of the participants provided clarifications to his answers in the final test, repeating one of his previous comments and providing two new ones.

7.3.3 Interviews

Corresponding to Marton's object of learning (2004), Winn's DPE framework and Laurillard's conversational framework (2001), here perceptions of both the theory and the toolkit are explored.

Perceptions of the CCO Framework

The study reconfirmed that the CCO framework on its own is difficult to grasp, because of its complexity and the effort needed to apply a general framework to a specific context. One participant suggested a formulation of what she identified as a possible obstacle. In the interview she phrased it as an open question: *"how to actually analyse the data in a form which can be useful… a crime is not a maths thing, which you can analyse, it's a big and complex thing. So all I thought is how you actually identify and find that useful information which can help you prevent the next time…."* (C06I)

Participants also made more focused comments about the particular aspects that were difficult. Many spoke of the ambiguity of the eleven generic causes. A few participants were critical of it, but there were two that appreciated the ambiguity in depth. One of them said the diagram *"looks a bit daunting. All these words you look at them and wonder what they mean. Some points are either too similar – I don't like that, but I can see the need for it. Or they seem too relevant, but I am not sure if I am giving the right information. It seems a bit overwhelming. I can see the need for it – if you need ten things, I have repeated some things because of that. The explanations are OK, if they weren't there I'd wanted them"* (E02D). The other commented on the balance of the number of rays talking that there is *"a lot of overlap, but not so much that any of these is redundant"* (E05D).

Many other participants went on to provide examples and recommendations of what did not make sense to them and how the diagram could be rearranged. For the purposes of this

research thesis it suffices to note that these comments are indications of the fact that after the use of the toolkit they were able to critically reflect on the CCO framework.

Two other themes that participants discussed were the difficulty of instantiating principles into new ideas and the challenge when participants had to think of the interconnectedness of those ideas. One of them said that *"brainstorming, new ideas – these were difficult"* (C14I), another one that they *"had to think about causes, effects, suggestions, ideas… this was very active. You had to do a lot of thinking and formulating for a technical subject"* (C25D). The interconnectedness of ideas was approached in the causal influences screen. That prompted one participant to comment that influences were *"too tedious"* (E27D). When talking of that screen another one said to have *"found [it] to be quite complex. And I guess it is complex because of all the interconnecting ideas… The layout itself is complex, because the ideas are complex…"* (C15D).

Misconceptions of the CCO Framework

There were a number of participants that explained difficulties in understanding that they still had despite having used the toolkit.

Talking of the framework one participant for example explained that the *"model applies to virtual.. does not apply enough to a real world"* (C15I). Then she elaborated *"in the online world the model makes perfect sense. In terms of actual reality people are more complicated. Why they do things, sometimes you need to go back to their childhood. Online stuff is all about information, presentation, transparency, the public persona of a company"* (C15I).

Another participant explained that she found explanations and terminology difficult to understand. Her further explanations showed her confusion: *"What I found difficult with other questions that followed was when it said that you have to come up with theories. So for me everything that was practical about the scenario and every time I had to give a practical solution or a practical explanation I found that easy because I could easily put myself in the scenario and imagine how things would work out. But then when I was trying to create theories, and trying to come up with abstract things again, then I wasn't really sure what to do."* (E18D)

Perceptions of the Toolkit

In this section I report participant comments that illustrate how they perceived the experience. The quotes provide evidence that participants were attracted to use the toolkit,

they were engaged while using it and they found something for them personally in it. All this goes to show that they did feel the beneficial effect of persuasive technology applied to the CCO framework. Some opinions of participants' about the way variation was implemented in the toolkit are discussed at the end of this section.

Various participants spoke of getting interested from different aspects of the toolkit. Some of these could clearly be mapped to Yee's (2005) motivational factors. A few participants shared that they were interested mostly to score better, like (E11), who stated that he tends to look at the final ranking only. Another one (E02) even described his situation as "*a special case, because I am the only person who was first and have it ranked… who had no motivation to do more things… enter more ideas.*" In an example of illustrating her case of *immersion* into the game mechanics of the toolkit (C06) explained "*I found the mechanics behind it interesting, the process of it.*" Another one discussed how she felt satisfaction from a specific feedback icon notification with the words "*it was quite encouraging when the new idea thing came up… So when I write something and it says "new" then I was quite encouraged and I would think of other points that would give me the new idea thing*." A participant (C06) also explained her preference for social interaction saying "*If there was more some way of interacting with other people. So that you can to discuss ideas… To have a dialogue back and forth between people, would make it a lot more interesting… Because if not, then it is just you and the computer. Which is not fun."* Another participant (E10) spoke of how seeing ideas of others made her think of more ideas herself: "*I think sometimes when I look at the comments and also when read other people I start to create new ideas, but I did not replicate the ideas."*

During the interviews several participants showed appreciation of the relevance of the scenario – something that arguably helped them engage with it. One participant said that it felt as a *"very live issue. And there's that companies are moving back and forth all over the world. This is a real time situation. So it is a very practical thing and that's why I took it so seriously"* (C04D). Another participant explained how widely applicable is the scenario problem. He was *"not necessarily looking at a scenario where you have a development company or development team, developing software, but in an everyday organisation, which is IT-centric, so they are using IT to take up their business. They are going to have an IT department. There is still opportunity for this kind of crime in any company"* (C01D). A third participant commented that it is an *"interesting case, because it shows how morale and employee treatment are intangible causes that lead to such*

tangible effects. That was a very interesting case… It sounds exaggerated, but I could foresee such cases happening probably to a lesser extent" (E08D). Yet another participant spoke of the value she saw in the examples: *"It was interesting, because it is big, important subject. You know, for personal users and businesses as well. So it did open my mind to all of the different forms of attack on computer systems"* (C25D).

Despite the widely acknowledged complexity of the framework, among participants there was also recognition of it being made accessible by the toolkit. In the words of one participant they gradually developed their understanding. He said that it was *"a lot of information coming at once. At first it is a bit overwhelming and after a while I got the gist of it."* (C26I)

Two main themes that could be identified are that the toolkit shaped the interaction of participants and that it helped participants focus on particular aspects of the problem while the toolkit ensured they could maintain coherence of the big picture. Regarding shaping interactions, one participant explained that he *"did it one by one. I looked at the different causes and then I went back to the scenario and looked what could have applied to that cause… of course [the causes] helped. Because it gets you to look at the scenario from a lot of different angles. Because you look at each cause, and then you go back to the scenario and look at it with that cause in mind"* (E05D). Another referred to the specific form this information was provided in, saying that *"it's good that [suggestions are in] a pop-up because once you finish this you don't need to see it again because you will be moving to another part."* (C25D). The last example also sets the other relevant theme here – letting participants focus on just parts of the solution at a time. Just after she had done that, one participant noticed that in her words *"I repeated a few of them… I didn't realise that I am repeating it, but when I saw them all together I thought that they have become irrelevant because I have already thought of them once before"* (C07I). All in all, one participant summarised that she found the task *"interesting, because it is clear and good. If you click on it there is clear explanation… [it] has a lot of information. It can be used by non-professionals"* (E19I).

Different participants explained what got them involved. Some participants spoke of the challenge for them to understand how the toolkit (and inherently the CCO framework) works. One of them said that she *"found the mechanics behind it interesting, the process of it"* (C15D). Another participant valued the challenge, saying: *"The enjoyment, it was interesting, seems very real. How can you prevent that, to deal with the employees? I*

was attracted also from a professional point of view." (E02D). One participant discussed even the actual intention of the toolkit: *"if the objective of the [toolkit] was to really make you think in detail about what things can affect the security of the place. Maybe it is good in the sense that it makes you think in a well-rounded view"* (C26D).

Participants also acknowledged that once they got involved with the toolkit, it also kept them engaged beyond just completing the task. One said: *"I did a lot of repetition. I was doing the task again, like being back here. I did it at home"* (E02D). Another explained his feelings about the opportunity to do so: *"I like the idea that I could get back home and then do it again, so there isn't much time pressure involved and I can do it in my own free time"* (E27D). Another one actually spoke of getting engaged in the topic beyond the toolkit *"I did a bit of research, had some friends in [the domain], checked out websites. It was something I always want to know"* (C04D).

A few participants shared that they felt encouraged by some persuasive elements. Talking of the new idea icon on the idea generation screens one participant explained *"Yes, and it was quite encouraging when the new idea [icon] came up. Because I'm not really good with this kind of things. So when I write something and it says "new" then I was quite encouraged and I would think of other points that would give me the new idea thing"* (E09D). Another one explained how competing involved her, elaborating on some of the related factors: *"I liked that there are three different [score] categories. So it gives a good idea of what exactly can you be the top in… if I know which is the best score I could try to work harder on it… The problem was there are some contestants who do the final boost. So towards maybe the last one hour they will type in all the ideas and overtake the first [competitor]"* (E27D).

Several participants acknowledged that the toolkit allowed them to have a distinctive personal take of the framework. This had to do with both personal interpretation of the scenario and choosing how exactly to use the toolkit.

In line with the first observation, several participants explained that they contributed, based on their personal experience. One explained it in her own words: *"my answers depended a lot on my background. I am working towards being a corporate lawyer. A couple of my answers are based on this background"* (E27D). Another found the scenario difficult and explained that *"I just knew it was something about IT there… [my ideas were based on] my own feeling and also the examples given in the rays"* (E09D). One participant elaborated that she found only some rays to be relevant: *"I was thinking about*

all of [the causes] and for me from the scenario there were just a few causes that could have created the problem. And I don't think that all those causes that the computer was offering were necessarily there in the scenario" (E18I).

There were similar reflections explaining selective involvement in providing comments in the assessment screens: *"when I rate [an idea] low I would try to give my comment to explain why I thought it was bad, but if I ranked it highly then I wouldn't bother commenting"* (E27D).

One participant (C03I) from the control group had ideas when variation could be useful. One one hand he mentioned that it might be good for learners that are new to the subject "*to see some some of the other people's suggestion."* He sees an opportunity to get a form of validation for the direction of thinking. A participant in the experimental group explains that it was more than validation for him, saying *"it was a bit [useful to see ideas of others], I got some more ideas"* (E02I).

On the other he discusses how variation relates to the common background of learners, saying "*it depends on the kind of people who are doing the study"* (C03I). Then he elaborates *"If we are all from the same school, doing the same kind of course, we probably all think the same way. So [seeing ideas of others] might be helpful."* He further doubts the scenario that might introduce wider variation of ideas, making the contrast *"…but if people are coming from different directions, I don't think it will be a good idea."* In this juxtaposition this participant relates to the comparison of narrow and broad variation, studied by Ranzijn (1991). As discussed in Section 2.1.1 the researcher found that wider variation is more challenging to learners, but makes more complex forms of knowledge accessible. In this case participant

Corresponding to Yee's analysis that social factors are more commonly a driver for female users, one participant explains how seeing ideas by others was more engaging for her saying "*because if [there are no ideas by other users], then it is just you and the computer. Which is not fun.*" (C06I). In this she implicitlymakes a reference to the richness that genuine multiple perspectives provide. In relation to this a participant (C03I) reasoned that "y*ou don't know whether the other person was right."* Yet, this is ambiguity that is typical for conditional knowledge related to interpersonal skills, and as suggested by researchers in threshold concepts (Atherton et al., 2008; Meyer and Land, 2006c) a challenge that learners need to get used to dealing with.

Frustration with the Toolkit

While some participants were positive about the experience, others were less enthusiastic. In this section we report critical opinions expressed. Although there were many specific recommendations regarding the usability of the toolkit, these are not reported here, being beyond the scope of this paper. Instead the focus here is rather on complaints that are inherent to the approach.

One aspect that was perceived negatively was realism. To one participant (C04D) the scenario had *"too short a period"* and *"too much stuff happening."* Another commented that in the scenario *"there's a lot of information. It's quite heavy"* (E02D).

A few participants (E05D), (E19D) commented that they did not know that comments in the score were written by someone else. This might have undermined participants' appreciation of the fact that they are actually engaging in a constructive dialogue about ideas. One participant commented that some of the ideas they had to assess were *"not realistic"* (E27D), which might allude to the fact that the toolkit used pattern-matching against previous ideas to provide immediate feedback, and in rare cases this lead to inaccurate attributions of ideas to comments.

Some participants were confused about what they should do. One of them explained *"to a certain extent unless if you were here, I wouldn't have been able to comprehend what I am suppose to do… If you weren't here and I only have these instructions I am not really certain I would be able to completely grasp the concept."* (C15D). Others were more specific saying that it was *"lengthy in terms of words."* (E27D).

Improved Awareness of Information Security and the Application of the CCO Framework

Several participants explained their (presumably newly developed) understanding of the complexities involved in information security. In the words of one of them he had a chance to *"realise that crime is not only about opportunity. It goes beyond what you see in the dictionary"* (C15I). Others went into more detail. One acknowledged that *"if a security manager fails… he's a crime promoter effectively, maybe not deliberately…"* (C17I). A few participants from the experimental group (as they were the ones subjected to it) acknowledged the benefits of peer-learning, for example explaining that the study *"It involved a lot of thinking and analysing techniques. I could see how some of the ideas other people have come up altogether later could be used to improve security methods and preventive techniques…"* (E24I).

Without this being explained to them previously, some participants also elaborated on how they understood why CCO makes a distinction between intervention principles and intervention methods. One explained that *"intervention principles are more general, and intervention methods are specific methods. A principle can have different methods, but they can come from a single principle"* (C06I). Another one explained that using the toolkit they can think of the different causes that one method can address *"[in interventions screen] I can only fill them each separately. But [in influences screen] a single intervention can have several influences at the same time. So here it is more [focused] on that what the different influences are."* (C06I).

7.4 Discussion

In about an hour the toolkit engaged participants in a structured discussion about insider threats and helped them develop a better understanding of the problem. In this session, they managed to come up with ideas for causes and interventions.

Overall, for many participants this session was enough to provoke deeper interest in information security and to help them develop relevant interpretations of the framework. This happened to the extent that participants could reason about and derive some parts of the work around the framework they had not been told about (like (E02) for example). As a result, and despite the fact that they had no previous experience with the CCO framework, the toolkit nevertheless allowed them to use the framework effectively.

The low number of remote participation did not allow for a conclusive statistical analysis as it was intended in the experimental design, or an in-depth study of the way learners engage with the toolkit. From the collected data, there are strong reasons to believe that the perceived complexity, inhibited many participants from engaging further. However, there is also a commonality with Study 1 (using *vLeader*), where there were a few exceptionally engaged students, a majority that engaged moderately and a few students that did not engage at all. In this study there were also participants that engaged at the bare minimum. These results were uniform across groups, so did not allow for conclusions about RQ3 based on behavioural evidence.

Yet, the exploratory interviews, reported in Section 7.3.3 in the part about perceptions of the toolkit, provided insights about the partially acknoledged benefits of variation to learning (RQ1) and engagement (RQ2). While participants appreciate the value of getting exposed to some variation, the challenges related to broader variation inhibits them to

appreciate how overcoming this difficulty allows them to develop deeper forms of understanding.

In providing a very competent explanation, some study participants, like (E02D), appreciated that this had to do with the actual interplay across causes and methods. The collected evidence demonstrates achievements in conveying the importance of social influence and threat appraisal (Siponen et al 2007) for insider attacks on information security. It also allows for a clearer interpretation of the specific opportunities to improve these further, especially when engaging with the issues of a particular organisation. One might speculate that if participants had had some information security experience, they might have found it easier to learn the CCO framework.

The expert assessment of intervention methods provide insights into the quality of ideas, suggested by participants. The low agreement between experts, and the high number of comments they provided suggests that many ideas are too vague and need further clarification. Once again, this was a result, common for both groups, so it did not allow me to draw direct conclusions about the research questions from the experimental setup. It also puts into question whether the toolkit managed to convey to participants the importance of being as specific as possible when suggesting methods. On the other hand, participants were not experts and the shortness of the scenario might not have given them enough context for them to be more specific.

Participants showed evidence that the toolkit (as much as they manage to perceive it separately from the framework) helped them to both get interested and understand the CCO framework better. The interviews reconfirmed that the employed motivational techniques were helpful. Participants acknowledged that the toolkit guided them to understand key aspects of the CCO framework better, but also saw it as challenging to work with. This effective persuasion demonstrated by the toolkit is an example of the application of security awareness techniques with high information quality, even if the presentation was only text-based.

Several sources of feedback reconfirm that the CCO framework is commonly perceived as too complex. There was a weak indication of the iGEQ question results, but also collected during the two subsequent interview sessions. This complexity confused several participants to the point that they misinterpreted what it was for. However, generally participants managed to grasp the essence of the framework and understand how to use it. The fact that they found the eleven generic cause categories vague and overlapping, did not

stop most of them from actually working with them.

The large number of cases when participants provided shorter answers after using the game-based toolkit, led to the consideration of several possible reasons for that behaviour. This was also an issue for the ideas for intervention methods that were assessed by experts. One obvious reason could be that they found that the essence of what has learned could be described with fewer words. Possibly they though that less information was sufficient to explain the concept. However, another reasonable assumption is that they experienced assessment fatigue and were less motivated to put effort into their second answer. A third potential reason that I identified is that they might have considered it unnecessary to repeat something that they had written before using the game-based toolkit not that long ago. This problem could also potentially be overcome by engaging with studies that would take participants through longer learning periods. Despite the fact that such studies require more effort to yield results, they might lead to more conclusive findings.

These results indicate that this approach of making information security frameworks accessible through persuasive technologies like the toolkit used here shows some early positive results. When given to ordinary employees at companies, generally it does improve their understanding of possible information security risks to the company and how they personally could either promote or prevent them. In combination with other research (Pahila 2007) this gives rise to the expectation that this will lead them to consider how to reduce their own contribution to possible risks and even engage them in taking the initiative on more extended risk prevention, for example by reporting noticed vulnerabilities to security staff.

These findings also gave me confidence to continue and take the toolkit out of the lab and further develop it for use in class, and ultimately in an actual organisation in the industry.

Furthermore, this toolkit could be adapted to the needs of organisations with bespoke scenarios that more closely correspond to the actual issues of the organisation and would be more relevant to its employees. Doing this to get employees to discuss suggestions will both allow for a new source of ideas, but also for better awareness about current insider threats.

7.5 Conclusion

This chapter described the lab evaluation of the toolkit developed in Chapter 6. Using an incremental approach to develop a serious game, I delivered a toolkit that was already

usable in a lab study. Frequently-iterated prototypes and usability feedback provided early signals on what needed changing in the game design and what could be kept. They also provided valuable insights into the underlying CCO theory and the process of teaching it. The user evaluations examined the types of engagement (as a contribution to learners' motivation to learn) and learning developed with the prototype.

The drop-out of participants during the study led to the decision not to split further participants in further studies into groups, unless the sample size is sufficiently big. This was not the case with the classes where the studies in Chapter 8 were conducted, so no design comparing groups was used there.

Engagement in the study was noticeably influenced by competition, discussion and feedback. These were not measured directly with a survey, but were a recurring topic discussed by participants. In terms of learning, the lab study provided evidence for better problem-solving, more creativity and critical thought.

Participants showed evidence that the toolkit helped them to both get interested and understand the CCO framework better. The interviews reconfirmed that the techniques borrowed from persuasive technology literature were helpful. Participants acknowledged that the toolkit guided them to understand key aspects of the CCO framework better, but also saw it as challenging to work with. This effective persuasion demonstrated by the toolkit is an example of security awareness techniques with high information quality.

7.5.1 Engagement

Overall, for many participants this session was enough to provoke deeper interest in information security and to help them develop relevant interpretations of the framework. These participants did not have previous experience with the CCO framework; nevertheless the toolkit allowed them to use it effectively. In the lab study participants exposed to variation were more willing to discuss the learning material by providing comments – a sign of higher engagement when exposed to variation in the learning content (which is the *intended object of study*). This is one form of evidence about the contribution of variation in the intended object of learning to the development of more engaging serious games.

There was a variation in the level of engagement between participants, that could be compared with Study 1 (using *vLeader*) from Chapter 5. In both cases only a portion of participants (around 20%) engaged with the game actively, and another portion of them that engaged superficially. This was seen in two different ways. Engaging beyond the

required minimum typically led to more idea generation and assessment during the initial period, but it also led to some form of engagement in the proposed competition.

Due to the low number of learners that used the toolkit remotely, I was unable to collect sufficient comparable data about long-term engagement and how variation affects it. Yet, during the interviews participants hinted that they felt motivational drives as in Yee's (2005) framework which can be seen as a call for better adaptation of the *lived object of learning*. The interviews also demonstrated how participant experiences during discussion and role-play shape and formulate verbally their *enacted of the object of learning*.

Even though it was intentionally build in as an engagement technique, enabling this personal level of variation was not part of the comparative study design. Yet, as one participant explained (E02I), this variation designed in the content played a role in their engagement.

7.5.2 Learning and Application

The learning assessment showed mixed results. All forms of assessment showed some improvement, but this learning could not be quantified to assess statistical significance, neither when comparing entry and exit tests, nor when comparing control and experimental groups. Whilst there is evidence that participants in the reported studies understood the framework and learned how to use the toolkit, there is insufficient evidence to conclude that participants were able to transfer that newly acquired knowledge to other problems in information security. The answers to the theoretical questions showed indications of the various forms of learning and corresponding varying evidence. While there were indications for improvement by the majority of participants, it was difficult to generalise these into distinctive common patterns for the whole group.

In the analysis of answers to theoretical questions, different assessment techniques aiming at different types of knowledge allowed for a comparison to be drawn between the forms of knowledge that participants developed with the game-based toolkit. It seems that participants were better able to apply their knowledge in context, than to formulate, explain or generalise it. Two reasons for this come to mind. One could be that they actually needed more time and broader perspectives to get a deeper understanding. Another possible explanation is that the toolkit is more suitable for developing procedural and conditional knowledge, rather than declarative, similar to problem-based learning techniques, assessed by Gijbels and colleagues (2005). Yet, this second interpretation was

not supported by significant improvement in the tests after using the toolkit.

A takeaway from the expert assessment of ideas for intervention methods could be that the toolkit needs to guide learners to be more specific in their ideas for interventions. This would also allow for better appreciation of the distinction between principles and methods.

In their answers to the theoretical questions of CCO most participants could explain the key features of the toolkit. I discuss the results, limitations of the study design and possible lessons to be learned from these.

Even though variation leads to better learning, in the lab study there was evidence that it also leads to higher variance in quality of ideas. One way to explain this effect is to think in terms of focus and self-confidence, which would still connect with Ranzijn's (1991) idea of encountering more complex forms of learning, and struggling with what is known as the liminal space in the literature of threshold concepts (Meyer and Land, 2006c). When learners do not have variation to confront them with alternatives, they stay focused on the one option that they have encountered (or invented). This allows them to dwell more on this particular option, but potentially inhibit their ability to see a bigger picture.

7.5.3 Study Method

I hypothesise that there are two major contributing reasons for the inconclusive results: shortness of the learning experience and imperfections of assessment. I develop an argument of the limited opportunity for engagement in the learning process that lab-based learning experiments allow for.

This chapter presents results of the development and continuous evaluation of a web-based toolkit. While it is useful to evaluate serious games in a lab setting in order to improve their usability, progress in learning might be more difficult to capture in a typical one-hour lab experiment session. Whereas lab-based studies are still necessary as formative assessment during the development phase of game-based learning tools, I suggest that class studies or longitudinal web-based studies are more appropriate to assess learning happening with their help.

Chapter 8: CCO Class Evaluation

After the successful completion of the lab study, CCO was also used in two consecutive class studies. This section describes the method and materials used for these studies and the results from them. The method and materials evolved between the first and the second study and this is described below.

8.1 Materials

Given that both the application of the CCO framework in information security and the toolkit are new developments, corresponding learning materials did not exist prior to the first of these studies. To overcome this four items were prepared. These included i) the toolkit website adjusted to the fit in the class; ii) presentation slides featuring introduction to CCO and a subsequent summary; iii) a paper summary of the relevant information; and iv) a student feedback form.

The following adjustments, not affecting the workflow (see Section 7.1.3 for description), were made to the website to make its presentation more relevant to the corresponding class. The introductory pages of the toolkit were adjusted to address class students and provide them with the relevant information in the context of the particular class. Registrations in the website were also adjusted so that in the score ranking students competed only with their peers, and not with all users who had previously used the system. Due to the early stage of development, in the first study the toolkit was referred to as "prototype", and in the second I decided to refer to it as "toolkit" in front of the students. In the first study a dedicated forum was set up as a way to provide online support, whereas in the second, such a forum was set up within the course management system (Moodle) that students routinely use in their class. The system in the second study was intentionally set up to allow anonymous posts. This was based on expectations, developed in the study reported in Section 7.1.2 that students might find it more comforting to ask questions without risking the personal embarrassment of publicly associating the question with their name.

The interactive toolkit was accompanied by traditional learning materials – presentation slides and printouts. The slides included a theoretical presentation of the CCO framework as it was taught to (conventional) crime science students. There were also two sets of examples: one previously developed case study in crime prevention and *worked out*

Question	Type
Lecture is insightful	5-point Likert-scale
Exercise is insightful	5-point Likert-scale
Prototype is insightful	5-point Likert-scale
Facilitation is insightful	5-point Likert-scale
Lecture is **attractive**	5-point Likert-scale
Exercise is attractive	5-point Likert-scale
Prototype is attractive	5-point Likert-scale
Facilitation is attractive	5-point Likert-scale
When considering the toolkit can you think of something that you found particularly **positive**?	open question
When considering the toolkit can you think of something that you found particularly **negative**?	open question
When considering the toolkit can you think of something that you found particularly **difficult**?	open question
Do you have any particular comments about the **scenario** screen?	open question
Do you have any particular comments about the **idea generation** screens?	open question
Do you have any particular comments about the **influences** screen?	open question
Do you have any particular comments about the **assessment** screen?	open question
Do you have any particular comments about the **feedback and score** screen?	open question
Was there anything that you found particularly difficult to understand?	open question
What do you think you learned during this exercise?	open question
Would you say that anything in your understanding of the subject has changed?	open question

Table 24: Questions in the survey administred after the end of the class activity.

examples for one of the information security scenarios (see Scenario *Meltdown* in Appendix C). Finally there was one slide that offered a summary of what were the long-term implications of what was studied, and of the *CCO toolkit* in particular.

In the first class a printout of the scenario was provided. In the second class students were given both scenarios (Meltdown and Spoonlure) and a printout containing the CCO diagram as it appears in the idea generation screens and the textual descriptions that emerge when any of the rays is clicked on. This was done in an effort to make the CCO framework even more transparent, despite the information being available in the software toolkit.

A feedback form (see Table 24) intended to capture immediate impressions was prepared for handing out at the end of the use of the toolkit in class. This included four sections. The first, quantitative, section consisted of eight Likert-scale questions. The questions asked

how “insightful” and “attractive” were any of the following four: lecture, exercise (addressing the tutorial class), toolkit (respectively “prototype” in first lecture) and facilitation. Provided possible answers featured 5 levels from “Not at all” to “Extremely”. In the other three sections of the form students were asked to provide more focused feedback about the prototype and the framework. Feedback on the prototype was separated into questions about context and interface. Questions about the concept were asking what was positive, negative or difficult about it. Questions about the interface were asking for focused comments on each particular screen. The framework-related questions asked about what was difficult, what students thought that they learned during the tutorial class, whether using the toolkit changed anything in their understanding of the subject.

8.2 Method

The toolkit was deployed in two consecutive classes at UCL titled People & Security, mostly attended by Masters-level students in Information Security. Compared to the environment where the *vLeader* study was conducted, the culture in this module allowed for more flexible involvement by students. This resulted in less clarity on how many of the enrolled students actually engage in class. As a consequence it was difficult to distinguish those that did not engage with the class overall from those that did not engage with the toolkit.

CCO was studied in a lecture and a subsequent tutorial (hands-on exercise). Online access to the toolkit was unrestricted, meaning that students could freely use it from the moment they became aware of it. This was meant to allow them scope for informal self-paced learning while at home. On a daily basis I reviewed user contributions and corrected feedback to new ideas that might have been misattributed by the system. This improvement included reclassifying misclassified ones and providing comments to innovative ideas (these are the ones that did not match any previously provided ones).

The lecture and tutorial provided two hours of face-to-face access to students over a week. When compared to the contact time used in the *vLeader* study, this was much more condensed and did not allow me to meaningfully administer tests before and after the activity. As a consequence these class studies did not include a direct measure of learning (an indirect one was featured in final class examinations, that I did not have access to).

In both studies interaction with students started one week before the lecture. The first contact was an announcement of what was planned to happen in the class and tutorial, and when. It was made on the course learning support system (Moodle) which was the main

form of communication. This announcement featured a hyperlink to the toolkit. During the lecture, the lecturer presented the theory and reintroduced the hyperlink to the toolkit. There was about a half-week gap between the lecture and class and in that time students were requested to use the toolkit at home. Our entire team of three (my supervisors and I) attended the tutorial and engaged in facilitation and discussion. At the end of the tutorial students were given the feedback forms to fill. They were also told to keep using the toolkit, and that they would keep receiving feedback on the responses they made in it.

The data collected during the two studies were contributions and logs, collected with the toolkit on one hand, and feedback provided in the dedicated forms on the other.

8.2.1 Class Study 1

The first class had 34 students were signed up on the course learning support system. The toolkit was in an early development stage, so it was recommended to students that they try it at home, but to form discussion groups during the tutorial itself.

Due to its early stage of development and the related uncertainties related to abusing the scoring mechanism of the toolkit, it was decided not to incorporate the prototype scores into the overall class assessment, as was suggested after Study 1 with *vLeader*. Instead, as a way to encourage students motivated by the pursuit of achievement, a prize draw was offered for the user that scored first according to any of the three (see previous section) metrics in the prototype – best overall score, best average score per suggested method and highest number of innovative intervention methods. Thus each metric, independently of the others, determined one winner.

The lecture presentation was 60 minutes. It featured a 30-minutes-long introduction to CCO by the framework's author, presenting it as it is being used in crime prevention, followed by 15-minutes example of how it was used in a real crime prevention case. Then, in the remaining 15 minutes, a tutorial through the toolkit was demonstrated with a class discussion of ideas (causes and methods) that could fit each of the 11 rays. Then students were briefed that they had the time to the next tutorial (about half week) to use the toolkit individually from home.

As a result of the limited engagement before the tutorial all students needed to be engaged more actively during the tutorial, which meant that they would have to generate ideas there and then, hopefully led by group members who had at least observed the prototype. In order to maintain the attractiveness of the prize draw, a decision was taken that two sets

of prizes would be given – one for ideas suggested remotely, and one for ideas suggested during the tutorial.

In order to make sure that any ambiguities were addressed, from the start it was planned that during the tutorial groups should use the toolkit from beginning to end, as had occured in the lab study. As an adaptation in response to the low student participation, before the start of the tutorial, it was intended that the first half of the hour would be used for idea generation, then 15 minutes for assessment within the toolkit, and 15 minutes for discussion of the results.

For the tutorial itself, they were asked to bring their own laptops or tablets and be prepared to work in groups of three students. Students were then invited to explore the toolkit. At the end their ideas would be jointly reviewed and assessed by the entire class.

At the end of the term, which was several weeks after the end of the study, but still as a concluding part of the class, students were given the presentation slide summarising the main learning points. This did not lead them to reuse the toolkit.

8.2.2 Class Study 2

In this second class study, students were given the opportunity to work on two separate scenarios. When logged-in, they were asked to choose a scenario they wanted to work on, and at any further point they could choose to switch to the other.

The class in which the second study was conducted had 61 students enrolled in the learning support system. As said, minor revisions had been made to the toolkit, but no changes in the workflow. This was accompanied by improvements in the process of using it.

In the second class no rewards were offered. Yet competition was inherent in the toolkit with each student receiving the three scores described in Section 8.3.1.

It was considered that there was insufficient time for students to use the toolkit in the tutorial itself, before having a useful discussion about their experience. This led to the decision to dedicate the entire tutorial to a class discussion without the use of the toolkit. Even though both tutorial discussions were facilitated by the entire team of three (my supervisors and I), this second one relied only on conventional learning materials (meaning it did not involve access to the toolkit). As a consequence a broader discussion of the scenario problems was possible.

At the end of the tutorial students were told that questions related to CCO would appear in the final-year exam. However, at the moment of submission of this thesis, exam results have not been released and thus were not included in this analysis.

Similarly to the previous class study students did not engage before the lecture and there was half a week between the lecture and the tutorial.

8.3 Results

In both studies only some of the students engaged with the toolkit. In the first study 8 out of 34 students used the toolkit. Three of these used the toolkit during the lecture, one used it between the lecture and the tutorial, and 4 used it at the tutorial for the first time. Of all that used it before the tutorial, only one did not use it also during class. This student has only registered in the toolkit without actually using it.

In the second class study 22 students engaged with the toolkit. As a consequence of the way it was introduced none used the toolkit during the lecture or tutorial. There were 18 students who used the toolkit in the gap between the two. Of them 8 used it again after the tutorial. A further 4 started using it after the end of the tutorial.

8.3.1 Involvement and Contributions

In a pattern consistent with Study 1 (using *vLeader*) and Study 3 (lab study with the *CCO toolkit*), there was a broad variation in involvement in the class studies as well. As in previous studies some of the students did not engage with the toolkit at all.

Among those that engaged with the toolkit, there was a broad variance in the level of engagement. This can be seen in Figure 23, which could be compared to Figure 9.

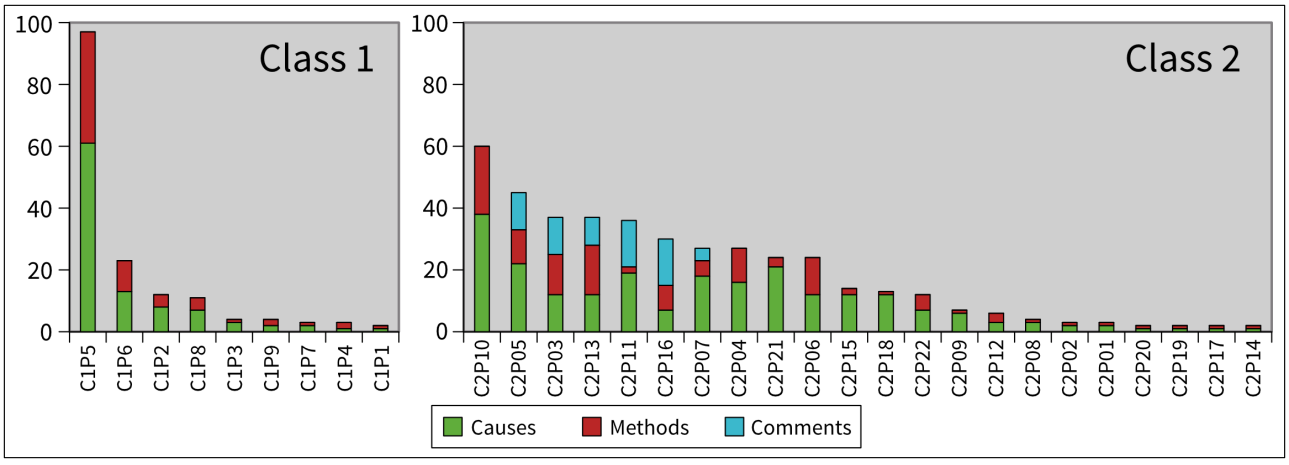


Figure 23: Ideas generated by students in each class as a measure of their engagement.

Class Study 1

In the first class study students were expected to explore the toolkit individually at their own pace before the tutorial. Two students did that during class and only one actually worked with the toolkit remotely. A total of 98 causes, 61 intervention methods and 48 comments were collected from three different users. The one that engaged remotely – assigned pseudonym (C1P1) for the purposes of the study – dominated contributed ideas. He alone authored 61 causes and 52 methods. He did not write any comments in his assessment.

Despite this adaptation, the tutorial did not run smoothly and thus undermined the learning experience. Several students were late (one arriving 20 minutes after the start of the tutorial) which resulted in difficulties organising the groups. One difficulty was that some students were waiting for their preferred peers to join them. In other cases later arrivals had to be told what has happened so far and needed to be included in an already existing group. All this led to a divergence of the speed of different groups, and close to 40 minutes spend on idea generation. The remaining time did not allow for a detailed discussion and the planned activities were only minimally performed.

A total of 32 causes and 17 intervention methods were generated by 6 users (groups), registered in the toolkit during class. Three of them wrote 27 comments responding to all methods they were asked to assess. The remaining users did not write any comments.

Class Study 2

In the second study 20 students explored the toolkit before the tutorial, and 16 of them actually contributed with ideas. They came up with 213 causes, 114 methods and 67

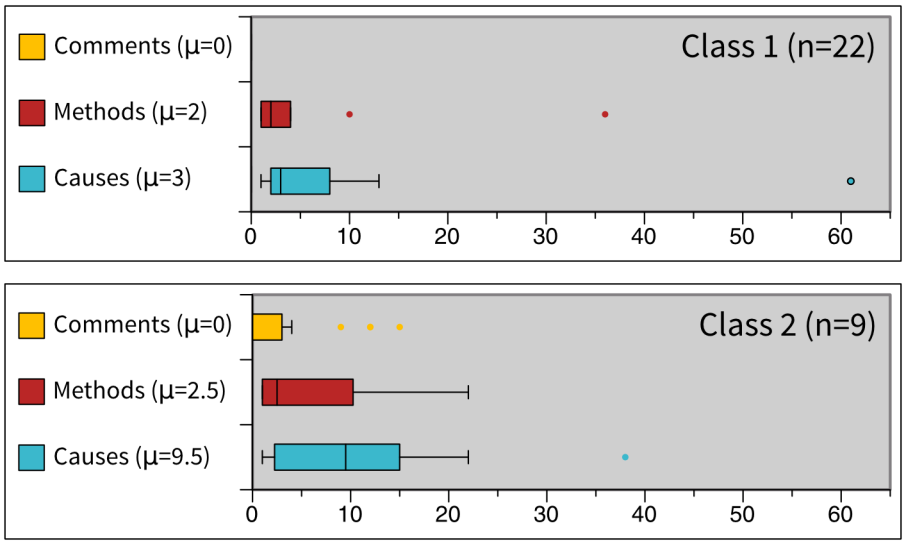


Figure 24: A box plot showing the number of user contributions in the two class studies. μ indicates median.

comments. In contrast to the first study, this was more balanced, as can be seen in Figure 24.

Without the pressure to use the toolkit in class, the discussion in the tutorial was much smoother and more efficient.

After the tutorial a further 7 explored the toolkit, but only 2 of them contributed with ideas. A further 15 causes and 8 methods were provided by these students. They did not provide any additional comments.

8.3.2 Feedback Forms

The results from the student feedback forms provided student reflections of the learning experience. A total of 12 feedback forms were collected in the first study and 11 in the second. Figure 25 shows a comparison between the distributions of responses to each of the 8 questions in both studies. As can be seen from the graphs for the first class, average responses dominated with all questions averaging between 2.7 and 3.4 and with median of 3 across all 8 measures. The averages for lecture and prototype were 3 or more and tutorial and facilitation getting averages up to 2.75.

In the second class all responses were noticeably better. All medians were 4 and all averages were 3.4 or more. There was clearly higher appreciation for insightfulness (which

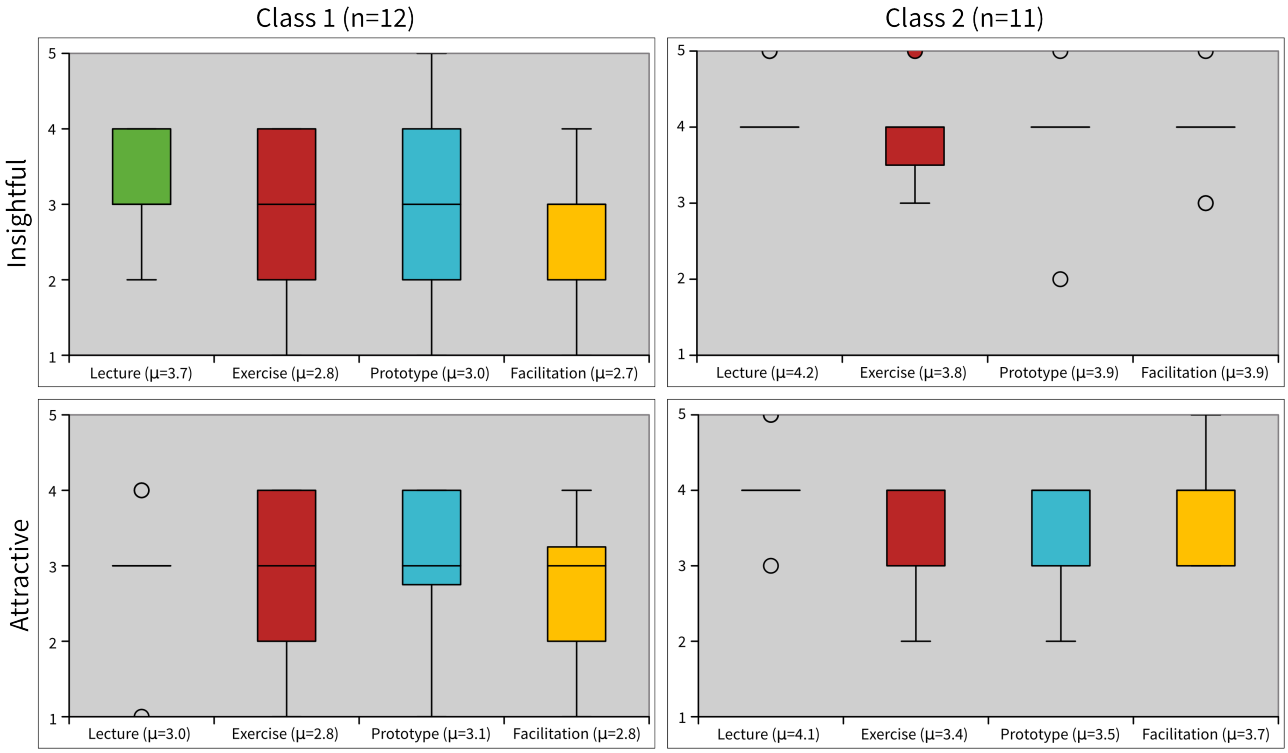


Figure 25: Student responses to the class survey after the use of CCO. The left chart shows how engaging students found the components of the experience and the right one shows how insightful it was. Note: *Here μ denotes arithmetic mean, not median. The median is uniformly 3 for Class 1 and 4 for Class 2.*

Construct	Scale	df	p(T<=t)	t Critical
Insightful	Lecture	20	0.0013	1.7247
	Tutorial	21	0.0057	1.7207
	Toolkit	21	0.0161	1.7207
	Facilitation	21	0.0001	1.7207
Attractive	Lecture	20	0.0005	1.7247
	Tutorial	20	0.0754	1.7247
	Toolkit	21	0.1068	1.7207
	Facilitation	21	0.0051	1.7207

Table 25: t-test with one-tail. The difference in df reflects the missing responses.

is how informative the learning activity was). There was also a strong increase in the attractiveness of lecture and facilitation, with tutorial and toolkit being somewhat less appreciated. As can be seen in Table 25 this also resulted in insignificant difference when results from the two classes are compared for attractiveness of the tutorial and the toolkit.

The answers to the open questions in both classes reconfirmed some of the observations in the lab study. When asked about what they found positive, students wrote that the toolkit is "*very intuitive*" (C1S3) and "*easy to use*" (C1S4). Others pointed out that it is "*very attractive and colourful*" (C1S1) and that they liked the "*design*" (C1S12). In the second class one student (C2S3) called the toolkit "*amazing, especially when it marks an idea as innovative*". They also wrote that it was positive that this was "*practical exercise*" (C1S2) and that "*there was a lot to think about, very engaging*" (C1S1). Two students in the second class (C2S2), (C2S6) liked the toolkit's structured approach. One of them summarised: "*well structured, understandable, educating*" (C2S6). In the second class two students (C2S8), (C2S10) noted that they appreciated the feedback they received.

When it came to discussing the downsides of the toolkit, again there was some agreement that it was "*too vague*" (C1S1), (C1S2), "*somewhat confusing*" (C1S6), "*very hard to apply the results in general way*" (C1S8), "*confusing questions*" (C1S11), "*too rigid*" (C2S10), but there was at least some appreciation of the benefits related to this ambiguity, one of the answers being clarified with "*this made us think out of the box so it was good*" (C1S1). One student (C2S8) criticised the duration of engagement of the toolkit saying "*The process seems too long for one sitting*" (C2S8).

Again, similar to recruited participants, students found that it was difficult "*to distinguish between some of the cases such as environment and enclosure*" (C1S1); "*understanding*

and applying case-relevant causes/influences" (C1S8), to "*come up with ideas*" (C1S12); and "*language*" (C1S10). Students did not provide much feedback to the other questions, except a few cases of reconfirming the same conclusions "*the differences between some (similar) categories*" were difficult (C1S7), "*don't understand how scoring is done*" (C1S4), and something also encountered in the lab study: the scenario would have been presented better with "*video and audio instead of text*" (C1S12).

8.4 Discussion

The class studies demonstrated complexities related to real-world research. This included a combination of factors undermining student engagement with the task. It is typical for students that they are assessment-focused (Biggs and Tang, 2007) and thus not very engaged for activities that do not count towards their final grade. The proportion of non-engaging students in these classes was clearly larger than in the case of *vLeader*. However, these comparison could be possibly related to the general tendency towards more flexible engagement in the module of these studies.

Also, the inadequate expectations and the subsequent ambitious plan for the engagement of students led to an unrealistic agenda for the first tutorial. This led to a weakness in facilitation that undermined the entire use of the toolkit. Quantitative results from the first feedback survey clearly show that students did acknowledge the problems with facilitation during the tutorial.

Feedback clearly reflects the fact that the second class went better. In the survey responses appreciation for the lecture and facilitation outperformed appreciation for tutorial and toolkit. Even though it could be considered that this indicates that the toolkit was not as attractive as intended, it does also show overall appreciation for the learning process, which is related to the intended objective about engagement.

As can be seen from the results comments on the toolkit were very similar to the ones collected in the lab study. In both cases there were comments about the general attractiveness of the toolkit, but also the potential to improve it further with more multimedia. There was also the discussion of the vagueness of rays, with some learners appreciating the need for such vagueness. This theme merged into the theme about the difficulty to contextualise the general guidelines of CCO into causes and interventions for the specific problem being analysed. Learners were complaining of an unintelligible scoring mechanism and challenging professional terminology from crime prevention

(students were more familiar with information security jargon).

Despite the suboptimal delivery of both the toolkit and the way it was delivered to students (facilitation), students engaged in discussion and successful idea generation. Although both the CCO framework and the toolkit are complex and difficult to understand, from comparing the two classes it becomes apparent that better facilitation significantly improves the experience. One particular change that was adopted in facilitation was that students were driven to engage in more personal discussion about the problem, rather than discussing the use of the toolkit itself.

8.5 Conclusion

This chapter described the development and evaluation work conducted to develop an early version of a web-based prototype. In the class study insights were collected about a number of tacit skills related to facilitation that are critical to computer-supported learning.

Consistently in these studies, similar to Study 1 (using *vLeader*) reported in Chapter 6, some engaged with the game for longer periods, and others engaged minimally, regardless of financial incentives (in the case of Study 4 – first class study using the *CCO toolkit*). Engaging beyond the required minimum typically led to more interaction (and in the case of CCO data generation) than during the initial period.

Despite its existence for over a decade and positive reception among crime prevention practitioners across the world, feedback collected in these studies, challenged CCO itself. Participants had very specific comments regarding the visual representation of the theory and the terminology used. As a result, efforts have started towards re-phrasing and re-designing the CCO diagram and its terms, but this is beyond the scope of this thesis.

Using an incremental approach to develop a serious game, I succeeded in delivering a mature prototype that already gives early results – both in terms of engagement and learning. Frequent prototypes and usability feedback provided early signals on what needed changing in the game design and what could be kept. They also provided valuable insights into the underlying CCO theory and the process of teaching it. The user evaluations examined the types of engagement (as a contribution to learners' motivation to learn) and learning developed with the prototype.

Chapter 9: Using Open Questions to Measure Deep Learning that Was Developed with Serious Games

Design and development of measures for assessment of deep learning when evaluating serious games presents a challenge of its own. This involves measuring both comprehension of new concepts and better understanding of the interdependencies of previously accessible knowledge. This chapter considers measuring deep knowledge for the purposes of evaluating learning technologies.

In the studies conducted for this thesis (and already reported in previous chapters) starting with the intended learning outcomes in mind, I aimed to identify evidence for the development of deep learning and different types of knowledge when learning with serious games. To do this in each of the two Studies 1 and 3 (respectively with *vLeader* and the lab study with the *CCO toolkit*) a combination of open-answer questions probing different types of knowledge was administered before and after introducing the corresponding serious games to the participants.

9.1 Method

The theoretical questions were qualitatively analysed using a set of analytical techniques. I used content (Weber, 1990) and thematic analysis (Aronson, 1994) to measure comprehension of new concepts, and the SOLO taxonomy to measure the interpretation of interdependencies of knowledge in the open-answer questions. As units of analysis, I used words and phrases for content analysis. Naturally, the unit of analysis was themes for thematic analysis and the application of the SOLO taxonomy.

Here I use "theme" to refer to what is called theme in thematic analysis and what Biggs called cognitive structures in the SOLO taxonomy. In cases when these are parts of the same topic, I sometimes refer to themes as aspects. This chapter includes a comparison and overview of the evidence of deep learning in these results that are illustrated with examples. Four types of learning are exemplified across both studies and discussed. These are explained below.

A typical incremental form of learning is the acquisition of new knowledge. It broadens the learner's understanding and allows for new perspectives towards the considered topic. However, simply counting the different aspects that learners enumerate is not a valid assessment, because one learner could list many basic aspects that are of little relevance to

the asked question, and another one could list only two or three that are more relevant. A typical example of such incremental knowledge is declarative knowledge, where truthful statements remain truthful independently of awareness of other aspects.

Another form of learning is developing one's understanding of the relationships and interdependencies between different aspects of the topic. This is exemplified by conditional knowledge, where awareness of one aspect determines how another one is being understood. In an extreme case the two aspects could fuse into a more complicated one, as represented by the two highest levels of the SOLO taxonomy.

This distinction is being addressed both in assessment with concept maps and the SOLO taxonomy. In concept mapping incremental knowledge is represented by concept nodes and relationships between concepts are represented by links. This allowed analysis to consider both types of knowledge and the actual structure of links (Hay and Kinchin, 2008). However, assessment with concept maps requires that learners are familiar with the concept mapping technique itself. The SOLO taxonomy is used to analyse conventional written assessment with open questions and considers the level of integration of concepts, thus determining the level of learning (Biggs and Collis, 1982). As a result, by analysing the themes present in a pair of answers (the ones before and after the learning task), both the number of themes and the connections between those themes can be assessed. However, when it comes to examining the themes themselves, thematic analysis allows for broader interpretation based on the learning context.

A variation of the development of links between concepts is explaining what was learned with one's own words. This is a demonstration that the learners understood what was learned and were able to represent it using concepts or vocabulary that they are familiar with. Such changes are of central interest in thematic analysis. The themes that might not have been taught or discussed in the learning activity itself, might be referenced in the learners' answers. Cases like these are beyond the scope of content analysis and the SOLO taxonomy.

A part of professional learning is the learner developing the identity of a professional, as described by researchers in communities of practice (Lave and Wenger, 1991). Part of this identity is the appropriate use of professional language. However, learning the vocabulary is relatively easy when compared to learning the actual expertise (meaning that using terminology could be an indication of a simple form of surface learning), and does not always allow for in-depth interpretation of the nuances of the used terms. Thus, whenever

Code	Question	Study
Q1	How do you understand leadership? What example could you name?	S1
Q2	You are working together with two other colleagues on a project. One of these colleagues regularly skips meetings and seems distracted when you discuss the project. His work has been of poor quality. What would you do?	S1
Q3	You work in a start-up with two other people. A major client has delayed a planned payment and as a result the company in turn cannot make a payment due by the end of the day. Your colleagues start an argument whether the company should stop working with that client. What would you do?	S1
TQ1	How would you describe the CCO framework?	S3
TQ2	What is the CCO framework used for?	S3

Table 26: Questions from previous studies aggregated in this analysis.

a learner has used professional language, the challenge is to identify whether the terminology was used appropriately. In other words, to distinguish whether the learner is thinking as a professional, or is simply imitating the vocabulary of a professional. This is where content analysis can detect the use of professional language, and thematic analysis would allow for discussion whether words are genuinely used or are simply a form of mimicry. Mimicry is a particular instance of rote learning (Hay and Kinchin, 2008) where professional language is imitated, but the actual meaning of the jargon is not intended.

9.2 Materials

Here I report the process that was taken to develop assessment materials for the evaluation of two serious game platforms in complex knowledge domains: leadership and negotiation in project management and crime prevention. The two studies were correspondingly Studies 1 and 3, reported in Chapters 5 and 7, respectively. Responses from these studies are compared here to reinforce the argument.

Study 1 was the only one conducted in a class that was included in this analysis. In it 47 students completed both the pre- and post-test written assessment, including the three open questions (Q1-Q3 shown in Table 26). In Study 3 answers were collected from 28 lab participants. The measures included two theoretical open-answer (TQ1 and TQ2) questions to explore participants' understanding. The study contained three further open-answer questions, but they asked participants to list ideas, which coincided with the way knowledge is applied in the toolkit. Such application questions lead to structured answers that are expected to be examples of the SOLO multistructural level (meaning that each item in the list provides a distinct theme). Hence, they are not included in this analysis.

The two studies included a total of 75 pairs of pre- and post-study short written responses of learners that were collected and analysed. The results from Studies 4 and 5 (see Chapter 8 describing the study) are not reported here. Due to the time constraint during the class activity, learning assessments were not administered before and after learning.

In Study 1 students filled paper test sheets which led to difficulties related to unintelligible handwriting. In Study 3 a web survey platform was used to individually present questions to participants and collect responses. In both studies at the pre-test, participants were given a first attempt at answering the questions to establish a baseline before learning. After playing the game, participants were asked the same questions again in order to probe how the learning experience changed their thinking and interpretation. They were not reminded what their earlier response was.

9.3 Results

Results from both studies revealed a number of indications of deep learning. These indications are typically diverse and are not captured by a single type of analysis that I tried throughout my research.

I present six stereotypical cases from this data that illustrate where different methods succeeded or failed to capture indications of learning. These six cases are: change in considering new aspects, change in the level of reasoning, relating to previous knowledge, adopting vocabulary, simplifying explanations and mimicry. These could be distributed into three groups – thematic signs of incremental learning (new themes) and relational learning (developing the links between themes); language signs of own phrasing and ambiguous answers that allow for multiple interpretations.

Consideration of new themes and level of reasoning are both captured by the SOLO taxonomy. Typically, adding a new theme to the answer indicated attainment of at least the multistructural level. Table 27 provides examples from each of the studies about each of the SOLO levels. The focus of analysis is on the answers after the task, but answers before it provide a comparison with the learner's understanding prior to the task.

As explained and exemplified by Biggs and Tang (2007, pp. 77–78), unistructural answers typically "*miss the point or… use tautology to cover lack of understanding. These responses can be quite sophisticated…, but, academically, they show little evidence of relevant learning*". From the examples in Table 27 (S1/P54/Q1) is of the former missing-the-point type and (S3/C13/TQ1) are of the latter tautology type.

Study/ n/ Question	Type of Evidence	Before	After
S1/ P54/ Q1	SOLO pre-structural	*I can understand the word "leadership" from any circumstances around me by notice myself or tought by my parents teachers or even friends in both direct and indirect way. For example I can see the role of my father who has leader role in my family.*	*I can learn to understand leadership from attending the lecture and class discussion, reading a number of articles involving in leadership details and also playing Vleader game.*
S3/ C13/ TQ1	SOLO pre-structural	*it is a framework which involve in cyber security and protect the end user.*	*secure and reliable. deter cyber crime from occurring. give people an idea about cyber-security.*
S1/ P02/ Q1	SOLO uni-structural	*Leadership is when a group of people follows the suggestions and ideas of an individual. For example in WWII Germany followed the leadership of Adolf Hitler which means leadership can be bad as well.*	*Leadership is the ability of influencing individuals in a way that your ideas are applied. In V-leader in order to achieve a good score you had to push through certain ideas.*
S3/ E11/ TQ1	SOLO uni-structural	*It tries to take a micro-approach in terms of identifying small problems in society that lead to a crime being committed (in terms of Information Security)*	*tries to reduce the risk and occurrence and severity of attacks by interrupting in the causes*
S1/ P17/ Q1	SOLO multi-structural	*Leadership is a characteristic which makes others turn to a person for guidance and advice. Leaders are charismatic and decisive, they keep the interest of 'the group' at heart. Garibaldi was a peoples' leader.*	*Leadership is the art of taking control by inspiring people through confidence and charisma. Leadership has to be emotionally intelligent because it is a two-way process between subordinates and superiors. Leadership tends to be associated with expertise in a given field. Inspiring & motivating & people focused leaders are the most liked among workers.*
S3/ C11/ TQ1	SOLO multi-structural	*CCO framework is used to reduce crime related to leakage or attack of information by investigating their causes.*	*it is a framework to identify the cause of crime related to information system followed by intervention to eliminate the crime.*
S1/ P41/ Q1	SOLO relational	*Leadership is a character of bring change to the organisation. People with leadership will try to find the opportunity of the business also infuence people in the company to work effectively.*	*I've understood from the lecture and the Vleader that leadership is the characteristic of being influential, motivating people to achieve their goals, solving the crisis in the organisation. It's about managing people, rather than tasks. Each scenario of the Vleader gave me some tactics of dealing with problems that likely to happen in the organisation.*
S3/ C01/ TQ2	SOLO relational	*CCO is used to identify current or potential breaches and to work through all the chains of effect, thus creating watertight solutions.*	*CCO is used to examine the many potential causes of incidents, and to explore what the implications of potential solutions would be, from all angles. Sometimes the implications are massive.*
S1/ P60/ Q2	SOLO extended abstract	*Depending on the severity of the situation I would talk to the colleague and try to find out the reason for his absense and poor quality of work. Thereafter, I come up with appropriate ways of dealing with him/her.*	*Understand as why he skips meetings and seems distractred. After which, depending on the circumstances and the reasons behind, appropriate set of actions will be taken. It's all down to the context the truth & motivation behind his behaviour.*

Table 27: Sample answers illustrating the different types of evidence according to the SOLO taxonomy.

The examples provided for unistructural answers demonstrate how these learners discussed one important aspect of the answer, but did not discuss others. In (S1/P02/Q1) the student discussed the importance of influence for leadership, but did not discuss anything about its situational nature for example – something that was focused on in the class and the *vLeader* simulation. In (S3/E11/TQ1) the participant talked of reducing attacks by addressing the causes, but does not go into explaining that this can be achieved through addressing those causes with interventions that get evaluated subsequently.

In contrast to prestructural and unistructural answers, multistructural ones addressed several aspects of the question. In the provided examples (S1/P17/Q1) discussed some leadership traits and leadership as a process; and (S3/C11/TQ1) mentioned causes and interventions. Although, clearly more sophisticated than unistructural responses, multistructural ones still are not guaranteed to provide the desired coverage of the question. They might fall short of explaining how the listed aspects come together to form topic of learning.

In relational answers respondents go beyond listing different aspects of the asked topic. In (S1/P41/Q1) the student talked of their expectations of how what was learned would be useful to them in the future: "*Each scenario of the Vleader gave me some tactics of dealing with problems that likely to happen in the organisation.*" In comparison, (S3/C01/TQ2) included a personal judgement "*Sometimes the implications are massive.*"

The examples selected to represent the extended abstract level are ones that describe similar personal judgement like relational answers. The main difference between the two is whether this judgement serves to inform decision-making in a more broader context, beyond the studied topic. In that vein (S1/P60/Q2) explained the process they would engage with in order to make a decision, thus exhibiting greater consideration of the context, than provided by the question. Unfortunately none of the 28 responses, part of Study 3 was identified to represent the extended abstract level of understanding according to the SOLO taxonomy.

Beyond the thematic signs, other signs in language could demonstrate certain forms of learning. Here expressing the newly experienced topic to one's own words; adopting the new professional language; and imitating the professional language, thus misusing it in mimicry are exemplified.

A few answers provided responses that used a language different from the one used in the

study, but still contained indications of learning. In one such answer (S1/P17/Q1) the student referred to emotional intelligence, "*two-way process between subordinates and superiors*", which do relate to topics discussed in class and in the context of *vLeader*, but were never formulated in this particular way. Similarly, in Study 3 one participant (S3/E14/TQ1) talked of "*precursors to crime*", "*going beyond the immediate causes of crime*", "*social causes like family upbringing, criminal motivations…*" which again are elements that are present in CCO, but in different wording and were definitely not part of the topic presentation for the study.

On other occasions respondents adopted the newly learned professional language in their responses. For example (S1/P53/Q3) used words like "*motivate*", "*situation*" and "*coaching*" in their response after the study, but not in the one before, thus starting to use these in the explanation after the learning experience. Similarly, (S3/E06/TQ1) included "*cyber attacks*" and "*effectiveness of such methods of intervention*," which were essentially present also in the previous answers, but utilising other wording.

Some respondents provided answers that used the professional language from the learning activity, but did not answer the actual question. For example (S1/P48/Q1) used "*trait, behaviours and reactions to situations*" in its wording, but did not actually answer the question. This is a representative example of a form of rote learning as defined by Hay and Kinchin (2008) – mimicry, or here in particular – using words without utilising them to represent their actual meaning. Similarly, (S3/C13/TQ2) talked of "*vulnerable people*" and "*cyber crimes*", but both fail to answer the questions in the specificity of their corresponding contexts. This is similar to prestructural answers as identified by the SOLO taxonomy, it is distinctive in the fact that words are used only superficially.

Beyond all these pieces of evidence, there were also a number of respondents that either simplified or repeated their answers in the test after the task. There are a number of different interpretations possible under these circumstances, but the opportunity to explore these in the interviews was missed.

One possible interpretation of simplified answers is that learners genuinely convinced themselves in the process of learning that it is simpler than they initially thought. In this case the shortened answer is an evidence of reconsideration and synthesis – a valuable process of reflective learning. Such a situation occurs when the two responses are phrased differently, and thus are not directly comparable. Rephrasing in this case could be interpreted that the learner does not engage in *mimicry* at a superficial level, but is able to

interpret and rephrase the answer. However, when there was no reformulation of the answer, such an interpretation cannot be made.

Other interpretations could be that due to a learning and/or assessment fatigue, where learners could not have been bothered to provide the level of detail that they provided initially. Table 28 shows the proportion of such answers in each of the two studies. These were distinctively higher in Study 3 when both pre- and post-assessments were conducted within a single session.

9.4 Conclusions

As a way to allow learners to express their own opinions, I used open questioning and examined learning assessment responses with a range of analytical techniques. Here results of the application of content analysis, thematic analysis and SOLO taxonomy are discussed and exemplified. Different forms of analysis showed different types of learning. The evidence provided here demonstrates the richness and multitude of manifestations of *deep learning*. Examples of answers that fall in the lower categories of the SOLO taxonomy show incremental knowledge like Ekblom's (2011a, p. 29) *know-what-works* and *know-when-to-act*, whereas higher levels show integration of knowledge (like Broudy's (1977) *know-with* and Ekblom's *know-why*). Even though content analysis and thematic analysis revealed different types of evidence, there is a similar line of distinction between adoption of professional language and explanation in own words. Effectively, all this contributes towards learners making sense of what was learned in the context of their own previous knowledge and way of thinking.

I have formulated and demonstrated successes and challenges for assessing deep learning with open questions across the studied domains. I have also examined limitations of the methods used. Finally, here I suggest potential alternatives to overcome these issues.

Assessment of learning is a task-specific challenge and assessment of deep learning is prone to be less formal and requires substantial effort to be conducted. It allows for indications of learning, but also for ambiguities.

Using free-form text-based assessment it is difficult to separate content from the form of its presentation. One issue is that some details might have been left implicit (contrast to concept maps, where concepts not included from the map are simply absent).

Analysing for short answers in lab sessions with the SOLO taxonomy presents further challenges, because of parts that are implicitly skipped, especially when having pre- and

Study/ n/ Question	Type of Evidence	Before	After
S1/ P17/ Q1	relating new to previous knowledge (e.g. own vocabulary)	*Leadership is a characteristic which makes others turn to a person for guidance and advice. Leaders are charismatic and decisive, they keep the interest of 'the group' at heart. Garibaldi was a peoples' leader.*	*Leadership is the art of taking control by inspiring people through confidence and charisma. Leadership has to be emotionally intelligent because it is a two-way process between subordinates and superiors. Leadership tends to be associated with expertise in a given field. Inspiring & motivating & people focused leaders are the most liked among workers.*
S3/ E14/ TQ1	relating new to previous knowledge (e.g. own vocabulary)	*The CCO framework goes beyond the actual symptoms of crime, and into scrutiny of the precursors behind the crimes. For example, these could be due to family background, or a market demand for black hat hackers. CCO goes into these underlying causes that motivate the criminal decision in Information Technology, and seeks to evaluate solutions with reference to their effects on these precursors.*	*The CCO framework works on precursors to crime, going beyond the immediate causes of crime. CCO looks at social causes like family upbringing, criminal motivations, risk/reward evaluations.*
S1/ P31/ Q2	adopting the vocabulary of what is learned	*First I would ask to that person what is happening as they are not working very well, I would ask if that person have a problem or why is difficult for them to come to the meeting. Once I understand the problem I have to think how to tackle the problem.*	*I need to think about how to motivate that person, ask them questions about is happening that does not allow them to focus and not deliver the expected quality of work. After understanding the situation better, I need to engage that person by different approaches and take maybe a coaching behaviour.*
S3/ E06/ TQ1	adopting the vocabulary of what is learned	*The CCO framework is a set of procedures that aims to tackle the problem of attacks on information security. The various procedures of this framework include, but not limited to, the identification of causes, intervening the causes and evaluating the effects of such interventions.*	*It is a framework that seeks to reduce the number of cyber attacks by identifying the causes of such attacks, intervening the causes to reduce the attacks, and evaluating the effectiveness of such methods of intervention.*
S1/ P48/ Q1	mimicry	*Mimicry is a characteristic given to someone within a role that they take on, whether it's within a company, at home or within a government. With leadership comes a lot of responsibility, the leader is in charge of the productivity of their group.*	*Leadership is not an innate quality that you are born with. It is a series of trait, behaviours and reactions to situations that add up to make different kinds of leaders. Some of these traits are harder to learn than others.*

Table 28: Sample answers illustrating other (than SOLO taxonomy) types of evidence.

Study/ n/ Question	Type of Evidence	Before	After
S3/ C13/ TQ2	mimicry	*it help people to reduce the chance of being cheated during online security.*	*protect vulnerable people from cyber crimes. give people an insight about increasing cyber crimes.*
S1/ P06/ Q2	simplifying explanations	*I would talk to my colleague about how his distracted work would affect the overall group performance and by how working together could lead to more productivity of the group. By attending the meetings regularly he would also be updated about the happenings of the group and would be prepared for any sudden tests.*	*I would discuss with my colleague about the importance of getting the project done on time & its effect on the overall grade. I will try to resolve the issue amicably.*
S6/ E14/ TQ3	simplifying explanations	*The CCO framework embeds crime prevention analysis within a broader social context, giving a richer and fuller picture to the analysis. It picks out the causes of crime, and evaluates solutions with reference to these broader, underlying causes.*	*The CCO framework puts crime analysis within a broader social context. It analyses how interventions affect other causes.*

post-tests with a short time gap between them (compare rates in Table 29). In all studies there are examples where participants mentioned a detail in the pre-test, but shortened their response in the post-tests, yet the share of such answers was much smaller when learning happened over a longer period. Reasons might be fatigue, or the assumption that it is understood (self-explanatory). Reminding answers that were previously given, made this more explicit, even if it did not really improve evidence of learning.

In the post-study test of Study 3 the question was asked again, without reminding the previous answer. Several respondents wrote exactly the same text, about 60 minutes later. This could mean that their thinking has not changed at all, or rather that the previous response was still fresh in their memory.

There were also cases where a respondent gave completely unrelated answers before the learning task and after it. This, in combination with the nature of the complex topics made it difficult to interpret what the participant had in mind. One possible interpretation is that this was a case of a change in response to reflection over what was learned. Another interpretation could be that it was meant as a complement to the original answer. A third possible interpretation is that it was simply a sign of disengagement, possibly due to learning and assessment fatigue.

To resolve ambiguities like the ones just described, educational research engages students in interviews about what was studied. Even though this was done as part of this research,

Study/ Question	Sample size	Total providing new evidence	Same answer provided	Less provided in answer
S1/Q1	46	87%	6.5%	6.5%
S1/Q2	46	78.3%	15.2%	6.5%
S1/Q3	43	79.1%	14.0%	7.0%
S3/TQ1	28	64.3%	10.7%	25.0%
S3/TQ2	28	32.1%	21.4%	46.4%

Table 29: Share of learners that amended their answers after the study. These are compared to the ones that provided the same answers as before, or only a part of it. In Study 4 answers before the study were reminded after it, so participants could only add.

the limited scope and focus of my research did not allow for extensive representative sampling of interviewees that would have led to conclusive resolutions of these questions.

Another observation from Study 3 is that deep learning can hardly develop in the 1-hour session on its own. In this session learners are presented the software and need to get accustomed to it. This makes engagement for longer periods a worthwhile objective.

When compared to the typically established assessment materials of class teaching, the assessment in this research was developed for the purposes of the prototype. This allowed us to only pilot it with a small number of participants, and thus questions have gone through a much shorter refinement process than ones in established learning activities. This was an issue in both studies.

The research effort in this chapter explored the internal validity of assessment with open questions. It does not, however, explore its external validity, that is how learners compare to each-other, because of the range of evidence of deep learning and the small samples of learners that actually exhibit each of them.

Chapter 10: Contributions to Research and Development of Serious Games

In this thesis I pursued contributions in three distinctive areas. These are i) the collection and analysis of evidence demonstrating the benefits of variation for learning conditional knowledge with serious games; ii) design-based research resulting in method and principles for the development of serious games to support conditional knowledge; and iii) a portfolio of assessment techniques featuring open questions to measure deep learning and conditional knowledge in particular. To arrive at this, I started working on this thesis with three research questions in mind, exploring how employing variation when learning with serious games contributes to improved understanding (RQ1), application (RQ2) and engagement (RQ3). This was moderated by the practical constraint that scientific evidence showed that different forms of interactive learning are beneficial to learning if they are addressing complex knowledge (Andersen et al., 2012; Gijbels et al., 2005). This constraint led to the consideration of learning subjects within disciplines that are sometimes called "soft" and are characterised by the subjectivity of the categories of right and wrong (Atherton et al., 2008). In the case of my research these were skills related to interactions involving people, represented by negotiation skills in the context of project management and the COO framework adapted to information security. Following from their complex nature, correspondingly adequate (Anderson and Lawton, 1992; Atherton, 2013b; Biggs, 1995) open questioning measures needed to be administered.

Research Goal	Findings
1. Study how can learning of conditional knowledge be supported by the introduction of variation in serious games	Development processes: - establishment of a framework leading from theory to game-based learning activities - evidence of different ways in which variation affects learning with serious games
2. Develop a serious game employing a design-based approach and refining design principles for serious games	Design principles: - the importance of being agile - feedback from implementation of the theory - particular needs for facilitation - reusing user-generated content
3. Develop assessment measures that capture deep learning and conditional knowledge in particular	- adaptation of the phenomenographic learning study - method for open questions assessment to measure deep learning as part of a portfolio of learning and engagement assessment tools

Table 30: Substantive and methodological contributions of this thesis for both research and practice. This table was introduced in Chapter 1.

In this final chapter (summarised in Table 30), I draw the conclusions from this thesis. I start with the substantive research findings from my comparative studies, along a discussion of the improved understanding of the process to conduct them and a more specific formulation of my research questions. Subsequently, I outlined the refined *design-based research* principles that are transferable to a broader range of application developments. Finally, the methodological contributions are summarised, first to evaluation research, and then to the practial development process. The chapter finishes with lessons learned and suggested development principles in serious game design accumulated in the process of evaluating existing games and developing the prototype used in the conducted research.

10.1 Substantive Outcomes – The Multitude of Variation

One key contribution of the research reported here indicates that the approach of making information security frameworks accessible through game-based learning showed specific positive results described below. I evaluated it with students, some of them part of a class, other as randomly recruited participants. Generally, they demonstrated engagement and improved understanding of possible information security risks and how they themselves could either promote or prevent these. Here I discuss to what extent these can be attributed to the variation techniques that were employed.

In its research on traditional classroom education variation theory explores aspects of the phenomenon being studied, or the *object of learning*, as in (Marton and Pang, 2006). In my effort to apply its principles to *serious games*, I reintroduced dimensions of variation similar to the ones identified by Pang, Lidner and Fraser (2006). To address the three identified manifestations of the object of learning, I have identified corresponding specific features. I have explored situational predisposition to learn or compete and *motivational*

Object of Learning	Understanding (RQ1)	Application (RQ2)	Engagement (RQ3)
Intended	Indicative evidence (Study 3)	Indicative evidence (Study 3)	Indicative evidence (Studies 3, 4 and 5)
Enacted	Indicative evidence (Studies 3, 4 and 5)	Indicative evidence (Study 3)	Indicative evidence (Study 3)
Lived	No evidence found	Strong evidence (Study 1)	Indicative evidence (Studies 1 and 3)

Table 31: Evidence of positive impact of different manifestations of the object of learning on the three research questions. Results from various studies conducted as part of this thesis.

factors to address the *lived object of learning*. The *intended object of learning* was naturally represented by variations in the learning content, the *enacted object of learning* was represented by *role-play* and *discussion*.

All these results are aggregated in Table 31 to provide an overview of what evidence was collected within this research and what needs further investigation.

10.1.1 Intended Object of Learning

When it comes to variation in the intended object of learning, as explained in the literature review previous research has shown how it is beneficial. This was difficult to explore with *vLeader*, without having access to change the actual learning content. Yet, as indicated in Section 5.4.1, the variation group outperformed other groups. This was further reconfirmed in the interviews after Study 3 (lab study with the *CCO tolkit*) where participants discussed how variation supported their learning. Even though it was not part of the experimental design, there was also the variation that the framework itself introduces (for example the 11 different rays). This led the majority of participants to engage with each of the provided alternatives and generate corresponding ideas. This showed that this particular form of variation very closely determined what participants actually did. Although the conducted studies did not provide an experimental condition when this particular form of variation was missing (something done previously in Study 1), there was observed evidence that generated ideas were direct responses of participants being exposed to variation.

10.1.2 Enacted Object of Learning

The enacted object of learning is what learners actually do throughout the learning activity and how they interpret it. This determines whether the learner's knowledge that was initially intended to be developed, will be conveyed accurately and correspondingly on how good the quality of the teaching process was. In order to explore their perspective (in addition to simply providing discussion opportunities for the sake of engagement) learners should be given an opportunity to verbalise their reflections and to discuss with others, thus exposing them to the variation introduced by the views of different individuals. This has two key benefits. On one hand, it allows learners to collect feedback on their ideas from discussions. On the other, it makes it possible to relate previous discussions to experiences that people are making later, thus allowing even players that play on their own to take part in an asynchronous discussion.

The value of this opportunity was almost universally appreciated throughout all interviews. In Study 1 (using *vLeader*) participants asked for the opportunity to generate own ideas. In Study 3 (lab study with the *CCO toolkit*) they appreciated the way it was implemented and asked for an opportunity to discuss even further.

Different people present different perspectives, but do not necessarily represent all viewpoints relevant to the scenario. This is why it is beneficial when a serious game drives players to look from the perspective of different roles, or in de Bono's words (1999) "to put a different thinking hat". As a technique this has been advocated for a variety of reasons (Ekblom, 2011b; Wills et al., 2010). This was present in both Study 1 and Study 3, as seen from the corresponding chapters here. Its value for both learning and engagement was appreciated by participants.

10.1.3 Lived Object of Learning

Variation in the lived object of learning addresses the fact that predisposition is one of the key factors to successful learning. This calls for efforts to try to introduce variation in the perception of the object of learning. In different studies reported in this thesis variation addressing motivational factors (feedback from Study 1 and design decisions in Study 2) was explored.

Due to different players (when using serious games this overlaps with learners) are motivated by different drives (Yee, 2005), a range of engagement techniques need to be embedded in the software to address individual motivational factors. In the research in this thesis I have used Yee's taxonomy of motivational factors (2005) developed in the context of massive multiplayer online games (MMOGs). Examples for such factors are the need to compete, the need to socialise or work in a group, the desire to find unexpected features and rewards.

In the development of CCO, special attention was put to address each of the broader motivational factors, so that the software could engage participants that are potentially driven by any of these. The features that resulted from these depended on the actual intended learning outcomes and the CCO framework (the object of learning). These led to decisions where competition could be introduced, what part of the content is optional (and thus concealed to be found only by the more persistent players), and where discussions could fit in place.

As a result of the specifics of the gamer jargon that Yee's survey used, some participants of

my studies had difficulties understanding the questions and I was not able to explicitly measure what attracted their interest. In participant interviews (for both *vLeader* and *CCO toolkit*) I indirectly explored whether participants felt motivated by features that could possibly correspond to any of Yee's drives. As can be seen from the quotes from Studies 1 and 3, participants not only acknowledged that these techniques were actually engaging, but also provided suggestions how to include even more corresponding engaging features.

10.2 Method Contributions

Two methodological contributions are made here. One of them – the adaptation of the *phenomenographical learning study* – had to do with transferring the original methodology from its typical formal educational environment to a mixed environment, where lab studies are used to allow for shorter iteration cycles. This is described in the first subsection below.

The third adaptation (described in Section 10.2.2) had to do with using open questioning as a measure of deep learning. This introduced a further level of difficulty, associated with qualitative assessment of deep learning, but is necessary in order to capture the more diverse and complex processes related to conditional knowledge and interpersonal skills in particular.

10.2.1 Adaptation of the Phenomenographic Learning Study

I developed a method for evaluation of learning that combined elements from the *phenomenographic learning study* and classical controlled experiments with *between-subjects design*. The purpose of this adaptation was to shorten the year-long learning cycle of the learning study and to adapt it for the purposes of lab experiments. As a result, I fitted two groups with control and experimental conditions within single iterations of my studies. This section reviews the specific challenges faced within such a setup.

The limitations of the studies I have conducted for this research thesis – the studies being cases of real-world research – are discussed and reflected upon here and ways to overcome them in future research are considered as a methodological contribution for research. Two limitations of this research had to do with the circumstances of the studies that were conducted within it. On one hand lab studies appeared to be too short for deeper learning, on the other the richness of learning assessment with open questions did not allow for a straightforward comparison between participants. Furthermore, there was a mismatch between what software was feasible to deliver as part of this research and what participants

expected to use when given a game, regardless that it is a serious game. Last, but not least, a number of research activities were considered to be beyond the scope of this research thesis and were maintained only to a sufficient minimum to allow for the performance of the critical tasks.

It is typical for lab studies that random participants are recruited for one or few individual sessions in a lab and are given remuneration for their involvement. As described in the corresponding method sections, this approach was also used in the learning evaluation studies reported here (Studies 1 and 3). However, one hour is far from sufficient for learning of complex knowledge, especially when learners have not proven their broader commitment at least in the form of enrolment in a wider class or degree. Furthermore, an established way to measure the outcomes of learning is by administering learning assessment tests before and after the learning task. As part of the portfolio of measures, I used open questioning, which demanded more effort to answer than typical MCQs. As a result, study participants had to put the effort to answer questions twice in a session (within 90 minutes in the case of Study 3). This emerged as one of the reasons for the very high number of responses that were ambiguous and did not show any improvement. This was an attempt to replicate a class learning environment within a lab study. Yet,the limited scope of participant involvement revealed the drawback of short time sessions: they were too short for either learning or assessment.

A desirable outcome of experimental studies is to validate the hypotheses with statistically significant differences between the control and experimental group. With the exception of the results about the effect of variation of support for the situational mindset has on application, I did not get statistically validated results. A key reason for this could be the small sizes of the cohorts I used, which was insufficient to measure more subtle effects that appeared to be weak in a short session.

The use of written open assessment technique has also introduced some downsides to this research. Even though the assessment method used allowed for good internal validity of assessment, but it is less beneficial when it comes to external validity. This is elaborated in the subsequent Section 10.2.2.

Although *vLeader* and the *CCO toolkit* formally comply to the definitions of a game, for example see (Sauvé et al., 2007), as the studies reported here show, often they were not perceived as such. Commonly players are not aware of definitions and have their own expectations of what a game should be. More often than not, these were driven by

awareness of contemporary entertainment games and resulting expectations. Participant comments about the need for graphics (or a more engaging media representation in general) rose issues of artwork and visual style. However, as it is discussed in Chapter 6, introducing professional art design was not feasible for the project.

The popular AAA entertainment games typically consume million-dollar budgets, and currently there is no business model of comparable scale in the educational technologies. This leaves researchers working with serious games to try to guide learners to develop different expectations when facing a serious game. In my case, this led me to present the used software initially as a prototype, and later as a toolkit in order to help learners develop the right expectations.

As it happens in real life, in each learning effort involving multiple learners, there is always very tedious performers that put enormous amounts of effort in learning (compare the leaders in engagement shown in Figure 9 for Study 1 and Figure 20 for Study 2) and as a result, bear the fruits of respectful learning success. There is also the few learners that for a variety of reasons do not engage with the learning material (conventional or technology-enhanced) and correspondingly never get the necessary involvement in learning. These reasons could be anything from personal negative attitudes towards the learned subject, to accidentally developing the feeling of discouragement by the learning materials or facilitators. For these extremes, one could argue that there might be factors beyond the researcher's control that countribute to such performances. As a result of this, the actual cases of interest are the majority of other ones, where these reasons mix into a difficult to untangle interaction. However, similar to Marton and Pang's necessary conditions of learning (2006), a first step is understanding the clearer extreme cases and then drilling into the perplexed ambiguous cases of learning. This research thesis has delivered exactly that first step. In doing so, it provided specific tangible contributions, but fell short of delivering the popularly expected experimental design with the objective learning assessment claimed to be used in the formal educational system. Yet, this is a problem that has long challenged educational researchers (Atherton, 2013a; Price et al., 2010) and progress is made slowly (Biggs, 1994; D. B. Hay et al., 2008).

There were a number of activities within this research which were necessary, but not critical to this research effort, and because of limitations in the scope, compromises were made. These include some aspects of the software implementation, as well as coding of data for the purposes of qualitative analysis. Yet this thesis established an adapted

methodology and reports early results that can be used for early guidance. These are used also to identify opportunities for further studies that need to be conducted with scientific rigour in order to claim conclusive results.

Coding of data is an essential part of qualitative analysis and is typically one point where subjectivity is introduced. To avoid this, it is typical in social sciences to verify the reproducibility of coding, by employing a second coder. Subsequently, an examination of the difference between the results of the two coders provides a measure of the validity of the method. In this thesis, this was done only for the practical part of the learning assessment of the Study 3.

In the other cases learner input (be it from lab study participants or class students) was analysed only by me, which allowed for experimentation with different analytical techniques (see analysis of results in Chapters 5, 7 and 8 for examples), but also leaves room for the above mentioned validation.

10.2.2 Analysis of Answers to Open Questions to Assess Deep Learning

As part of this research, two studies (Studies 1 and 3) involved learning assessment and correspondingly, open answer questioning. When considering also complete responses during pilots, this resulted in a total of 75 answers of participants analysed. Each of these were a pair of answers that participants provided before and after the learning activity. Based on the results of the written assessment, observations of commonalities in responses and conclusions for the assessment method could be drawn. In this section successes and challenges for assessing deep learning with open questions across the domains I studied are formulated, including limitations of the method. The range of domains suggested that those conclusions could possibly be generalisable to other areas that have complex knowledge involved.

The written open assessment described in the previous chapter allows for relatively good internal validity, as it captured different form of improvement in participants' understanding. I used content analysis, discourse analysis and the SOLO taxonomy. Content analysis was less useful for the lab study when participants came from different backgrounds, and did not engage with the material long enough to develop a wider change in thinking and particular common professional vocabulary. Discourse analysis allowed to better consider references to the learning material and participant's background that are made within the context. This allowed the analysis to capture when participants related

their answers to their previous knowledge and to consider the variance in the used vocabulary. The SOLO taxonomy provided a method to analyse the structure of answers by suggesting a way to compare structural evidence of learning, thus defining (at least) a partially ordered dimension of possible answers.

Open question assessment introduced further complexities. Comparison between answers to the assessment before and after engaging with learning, allowed for a discussion of internal validity, but not external – whether participants reached a particular type of thinking after learning. This was partially due to the specific for lab studies lack of common background of participants that was discussed above. However, it is also challenged by the nature of complex (or "soft") knowledge, where alternative approaches, suggested in the learners' answers, were difficult to declare incorrect.

Results from the two studies showed a number of indications of deep learning as suggested by different sources in the literature. These indications are typically diverse and could not be captured by a single type of analysis. The four stereotypical indications of learning explained in Chapter 9 and reiterated below, demonstrated the learning that took place with the games that I studied. I further demonstrated several stereotypical responses that exhibit different forms of ambiguity inherent in them. This did not allow for an assessment with sufficient confidence. Potential alternatives to overcome these issues are discussed in the final chapter.

The assessment conducted here identified four types of evidence of learning. These are change in complexity of reasoning, considering new concepts, relating new to previous knowledge and adopting the vocabulary of what is learned. However, there was also a multitude of answers that were ambiguous to analyse with these techniques. As a result of this, they were difficult to uniquely assign to any of the above classes of evidence of learning. The issue typically arose when a participant responded with a shorter answer after engaging with the serious game.

Another form of learning, typical for communities of practice (Lave and Wenger, 1991), is when learners start using the professional vocabulary of the topic they are learning. However, including this in the assessment criteria leads to the risk of tolerating *mimicry* – a typical form of *surface learning* where learners use the professional vocabulary, but do not understand its meaning in depth. In some particular cases because of the shortness of answers it was difficult to distinguish when an answer using professional vocabulary was the result of an analysis, and when it was simply formulating "something that sounded

smart" (a case of *mimicry*).

As Biggs discussed in a paper considering the relationship between knowledge and learning (1994), learning could both provide learners with new knowledge, or it could help them better understand the relationships of what they already know. Using concept maps as a form of visual assessment, Hay, Kinchin and Lygo-Baker (2008) enforced a structure where these two types of learning are clearly distinguishable – the former is being expressed as the inclusion of new concepts, and the latter is represented by the connections between concepts. The combination of forms of analysis used here, allowed for an interpretation of answers and the knowledge represented in them in a similar way, while not requiring from learners to get used to concept maps. Instead, written assessment is used, a technique learners are commonly more familiar with.

Clearly, changing the level of reasoning and relating the learned topic to previous knowledge are of the second structural type explained above, and considering new knowledge, or adopting new vocabulary are of the incremental first type.

The different types of learning lead to a personalised variety of learning outcomes, that are challenging to be generalised. Thus, in contrast to the single scale of the multiple-choice questions, the outcomes can only partially be compared, and can be mathematically represented as lattice at best. Compare this to the visual representation of concept maps (D. Hay et al., 2008) or the SOLO taxonomy diagram (Biggs, 1995) recreated in Figure 3.

At the same time there are different ways to express phenomena, esp. when talking of complex knowledge as was the case in the conducted studies. It is unrealistic to expect within such short learning procedure someone to capture all possible (different) perspectives that could explain the phenomenon, even more so within a short assessment session.

The effectiveness of this evaluation methodology was undermined by on of the practical constraints of this research effort. Comparison between assessment results of Study 1 and the other conducted studies, showed that the answers to open questioning in lab settings are less informative, when compared to class settings.

10.3 Design-based Research Principles

As a case of design-based research, this thesis serves as a case study for the advancement of principles that can guide effective development, implementing currently established good practices. As a consequence, I developed a framework guiding development and

collected evidence for the importance of four general principles. The first of these is the use of agile software development when working on learning technologies. The second is using the feedback that operationalisation into a software system provides back to the theory it is based on. The third principle is the provision of structured facilitation supporting the use of the software. The fourth and final principle is reusing user-generated content as a source of feedback to learners. Each of these, starting with the framework is discussed in further details below.

These principles are possibly of greater relevance to conditional knowledge where a unified and established theory is lacking and there are a number of emergent, possibly partial and sometimes contradicting theories.

10.3.1 Contribution to Development Research of Serious Games

The framework developed here is a design-based research methodological contribution that is related to the difficulty of developing a serious game that addresses a complex area of learning. Even though design-based research already proposes an iterative approach to development, games are a multifaceted piece of technology, which raise further challenges. As a result, a project of developing a game by a single developer – further strained by the evaluation and reflection necessary for quality research – requires very careful planning and efficient development. To this end, I have employed a development lifecycle consisting of exploring existing *game mechanics* (concept phase – Study 1); early testing of game usability and paper prototyping (design – Studies 2 and 3); and iterative refinement of the final product (production – Studies 4 and 5). An overview of the entire process that was adopted is provided in Table 32.

Each of the three phases provided critical insights for advancing the prototype further. The studies of the concept phase allowed for an exploration of existing game mechanics and to what extent they have the potential to meet the needs of the current development. The studies in the design phase allowed for paper prototyping initially, and later for rapid

Study Order	Phase	Type and purpose of studies
1	**Concept**	Proxy study to test the evaluation methodology before own game is available
2	**Design**	Formative small-scale lab evaluations to establish a fun game
3		Summative lab evaluation to validate learning and engagement
4	**Production**	Real-world studies with the mature product to establish adoption and facilitation process

Table 32: Development life-cycle for the purposes of design-based research. This table was introduced in Chapter 4.

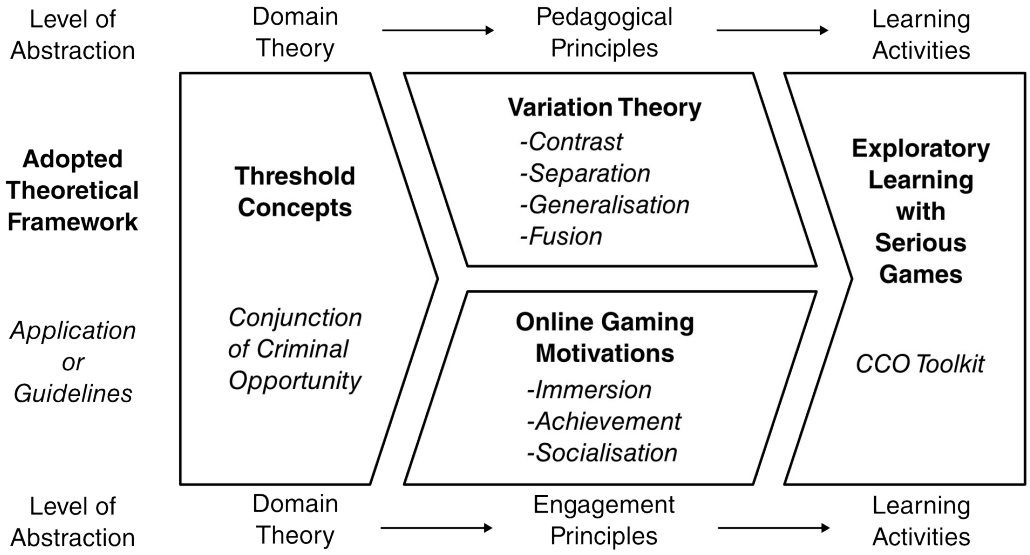


Figure 26: The process of deriving serious games from the theory intended to be taught. The figure first appeared in Chapter 1. The process starts from the theory being taught, continues with applied pedagogical and engagement principles, and finishes with the learning activity containing the CCO toolkit. Regular texts in the figure reflect the framework. In bold are the theoretical frameworks used in this thesis, and in italics are the particular instantiation here.

(software) prototyping. This in turn allowed for short development and feedback cycles until a version was reached that is both engaging and beneficial to learning – something that is particularly challenging, even if fundamental for serious games (Hays, 2005; Susi et al., 2007). Finally, the production stage studied how the software product can be embedded in the social environment and learning process of a university class. This provided the necessary feedback in terms of preparing supplementary materials and settling on a specific sequence of events, depending on the length of classes and time for self-paced learning students have between events.

On the other hand, the design itself consisted of applying pedagogical and engagement principles to the CCO framework, according to the model represented in Figure 26. This 4-part model describes how the relevant threshold concepts can be addressed with the design of a serious game, ensuring the inclusions of principles that address both learning and engagement. Here variation theory was used to guide the pedagogical principles, and Yee's motivations of play guided engagement princples.

10.3.2 Agile Software Development

Agile development and *rapid prototyping* in particular, is well established in software development as a way carry out iterative development. Yet, to my knowledge it has not been referred to in *design-based research* literature.

When working with complex knowledge and emerging theories, it is even more difficult to

establish how should the key intended learning objectives be addressed and what kind of experience would allow learners to master them (Cousin, 2008). As a way to overcome this in the development of the *CCO toolkit*, I used paper prototyping for its earliest usable versions and *rapid prototyping* with frequent formative evaluations during the software development process. This revealed which form of interaction would be more accessible for learners to engage with. It also served as early guidance in the development process, and determined the decisions related to the toolkit interface.

Having continuous release plans as part of the *rapid prototyping* approach, allowed for early and frequent trials and feedback from users throughout the process. Resulting from this approach, evaluation activities had to be adapted accordingly. It allowed for earlier insights into the implementation of certain design ideas and consequential early remedial actions in the game design. This was accomplished by initially focusing the collection of evaluation feedback on usability and motivation. Just then (when there is evidence of the tool being somewhat engaging), it became possible to explore learning through learning assessment.

This agile approach resembles the iterative development typical for design-based research. Still, there are two differences worth noting. The first one is that lab studies were added to intensify the feedback process, and the second – the initial feedback was focused on engagement, postponing learning feedback for stages when the prototype is considered engaging enough.

10.3.3 Validating the Theory and Feeding back to It

Despite its existence in over a decade and positive reception among crime prevention practitioners across the world, feedback collected in these studies challenged the CCO framework itself. Participants had very specific comments regarding the visual representation of the theory and the terminology used. As a result, efforts have started towards re-phrasing and re-designing the CCO diagram and its terms. Yet, this theoretic and knowledge-management process is beyond the scope of this thesis.

Feedback collected during Studies 2 to 5 about CCO provided documented evidence that the framework is commonly perceived as too complex. This complexity has confused one participant without prior awareness of security to the point that he misinterpreted what its purpose was. However, the vast majority of participants managed to grasp the structure and purpose of the framework and understand how to use it. The fact that they found the

eleven generic cause categories vague and overlapping, did not stop most of them from actually working with that ambiguity. Dealing with such uncertainty is an extremely important element of provisional learning. Some participants appreciated that this had to do with the actual interplay across causes and methods.

The collected evidence demonstrated achievements in conveying the importance of both *social influence* and *threat appraisal* as these have been identified by Pahnila, Siponen and Mahmood (2007). It also provided them with the necessary background knowledge to consider some other important aspects that they were not taught about, or to express criticism to the application of crime prevention terminology in the domain of information security. Finally, it allowed for a clearer interpretation of the specific opportunities to improve both theory and application further, especially when engaging with the issues of a particular organisation. It is possible that if participants had some information security experience, they might have found it easier to learn and use the CCO framework.

Providing this sort of feedback to the theory is one aspect of the teacher-researcher collaboration that design-based research involves. The second one is the importance of professional facilitation provided by a teacher, as described in the next subsection.

10.3.4 The Need for Facilitation

Feedback and reflection are crucial in learning, and the traditional way of making sure they happen is facilitation, often described as *scaffolding*. (Pang and Marton, 2007). Yet, facilitation plays a role in computer-supported learning (Kirschner et al., 2006; Sweller et al., 2007) that involves a number of tacit skills, and thus, is not something that could be passed along simply by documentation. This has been acknowledged in *design-based research*, but results from this thesis reinforce this argument.

Even when a game tutorial exists (as was the case with *vLeader*), learners need to be supported by a facilitator at their first hands-on encounter with the software. This is important for two reasons. One is that there might be ambiguities in the toolkit or its inclusion in the particular learning activity. The other reason is that possible discussions which might not have been anticipated during the development of the toolkit. In Study 1 (class study using *vLeader*) there were indications that although everyone learned, mostly as a result of the accompanying class, the students subjected to variation have adopted the domain jargon faster, followed by the control group that also outperformed the ones that did not play at all.

In Study 1 after the experience students were more inclined to describe leadership in terms of influence, which indicated how their way of thinking of the subject had come closer to that of researchers in the field, see for example in Fisher and Ury's work (1991). Although, this was a common trend between all the groups, it was stronger with those playing the game, which indicated that the game could have helped them to reconfirm what they had been taught in the class. In contrast, the *CCO toolkit* is not as mature, so it has even greater need for a facilitator to monitor and intervene on occasions.

Particular tacit facilitation skills that I identified are familiarity with the difficult aspects of the learning material and monitoring learners for early signs of loss of interest.

Learners do not explicitly communicate all of their intentions, thought process and mood to the software. This is the reason why continuous observation of the learners themselves, and signs of feedback that are currently infeasible to be captured by technologies need to be managed by a facilitator, who would be able to decide when to encourage learners to engage harder, or advice them to take a break. Examples of such intrinsic to the learner processes (that an experienced facilitator can sense subtle signs of) are learning fatigue and/or shifts of interest that need to be gauged in order to maintain the learner's involvement.

Arguably, familiarity with the learning material and tools where learners stumble upon difficulties are critical to such facilitation as suggested by research on threshold concepts, for example by Davies (2003). At least some part of this process can be externalised and formalised through methods like the Japanese *lesson study*, or its derivative phenomenographic learning study (Pang and Marton, 2003).

10.3.5 Reusing User-Generated Content

The analysis of the content collected with We the Giants showed that there is a number of recurring ideas in user contributions. This led to the attempt to try and identify such matches in CCO, and reuse the feedback that previous ideas received. The platform had an objective to encourage collection of focused user-generated ideas and comments on specific crime problems.

This is possible when the toolkit guides players to converge their contributions towards the objective of the game. As a result, participants commonly used words in a narrower context and adopted domain-specific language. The *CCO toolkit* focused contributions on solving given scenarios in crime prevention by applying the framework. Contributions were

automatically classified to capture repetitive ideas and cluster them in commonly recurring classes. This idea matching mechanism allowed for immediate feedback when players contributed, reusing the assessment and comments on previous similar contributions and higher chances of identification of innovative ones – ideas that do not match any previously existing class. Such innovation is particularly important in fields like crime prevention, because of the existing arms race between offenders and crime preventers (Ekblom, 1997).

It is typical for clustering data in natural language processing that it is an imperfect process and inevitably includes misinterpretations of player contributions. Still, because it is done as a form of learning support, players are expected to consider the feedback that they receive critically, and challenge it in a discussion.

Similar to what was carried out with the *CCO toolkit*, and taking place in parallel, Swanson and Jhala (2012) have independently developed another game that also collected learner's knowledge in a structured and reusable form in a complex knowledge domain. The process used in this research could possibly be transferred and adopted in other serious game development projects, leading to a low-cost development and early feedback on game design ideas.

10.4 Future Work

The research conducted as part of this thesis has allowed several previously implicit dimensions of learning to be explicitly formulated and some of them examined for their impact on learning. It has also suggested a combination of several text-analytical methods to allow for assessment of learning of complex knowledge, and has provided three case studies of using this combined method. Finally, it has tested and validated in specific context several practices. Based on this work, and out of the wide range of opened and unanswered research questions, three specific directions are discussed and are to be pursued in future. These are the continuation of the practical development work for this and other applications, generalisation for use in other learning contexts of the facilitator methodology developed in the production phase, and improvement of the study methodology so that the previously identified sources of noise in the assessment results can be reduced or mitigated.

In terms of continuation of the practical development, there are two envisaged directions: refinement of the development and validation with other research frameworks. As already

explained in the previous section, the work conducted for this research thesis entails a number of activities, which were considered not to be critical to the delivery of a self-sustained thesis itself, so were delivered at a minimal satisfactory level. This was done with certain and obvious compromises with scientific rigour and this needs to be corrected. As already explained, this includes developing software components that are currently only method stubs or mock objects, and revision of qualitative coding by a second researcher with a quantitative comparison of the results of the two pieces of coding to measure their level of agreeableness.

Another prospective continuation of the development, could explore how the adapted method (refer to Figure 26) could be applied to other research frameworks. For example Fogg's *captology* (2009, 2002) can be used to address engagement, instead of Yee's motivational factors (2005; 2012). Similarly, approaches from *cognitive load theory* (Artino, 2008; van Merriënboer and Sweller, 2005) can be used instead of variation theory's necessary conditions (Marton and Pang, 2006; Pang et al., 2006) as a way to allow for more accessible incremental approach to learning. Furthermore, Laurillard's *Conversational Framework* (2009, 2001), or Winn's Design-Play-Experience framework (Winn 2008) can be used to replace the manifestations of the *object of learning* as adopted from variation theory (Marton et al., 2004; Pang et al., 2006). These alternative research frameworks have broadly similar implications, but work through different set of principle recommendations, following which would lead to some different design decisions. It would be of interest to compare which framework is more accessible to implement, and hopefully to explore which could lead to greater benefits for learners.

Facilitator support during the use of the toolkit will be beneficial to better learning. To aid the process of facilitation, materials need to be prepared. A first round of learning materials were developed for the conducted studies. Still, a reflective process similar to the *lesson study* (Pang and Marton, 2003) could provide better insights in what particular aspects are difficult, how are they being perceived and how they could be better catered for. Such research will also contribute to the understanding of potential threshold concepts in the field.

The third and final direction of future work is further investigation on the effects different forms of variation have on learning and motivation. Whenever possible and when the toolkit is mature enough, it is preferable to conduct these studies in a class environment. Whenever appropriate classes cannot be identified within a reasonable timeframe, online

controlled experiments could be conducted. Each of these is discussed in the following paragraphs.

Beyond the possible immediate continuations of this work, gaps in parallel research have also been identified. Most importantly, this is the adaptation of the CCO framework for the purposes of information security. Despite the very broad recognition of the framework's usefulness for the purposes of information security, there are also strong indications that the language used is confusing for users in the new domain. Also, the 11 causal rays of CCO (see Figure 12) might need to be revisited as a way to clarify how terms like *enclosure*, *presence* or *target* translate in cyberspace. Although some interpretations were considered both in my team of researchers and by participants, a broader scientific inquiry might be useful in the process of clarifying and converging these interpretations.

The methodology developed and used in this thesis is also of interest on its own. As already discussed class studies engage learners for longer periods (for example one week, as in Studies 4 and 5), so they provide a better learning opportunity (environment).

Another major difference between class studies and lab studies is the motivation of participants. As a contrast, in lab studies, typically people attend because of the financial incentive, and the toolkit aims to genuinely engage them further. In class studies financial incentives are not possible and only some learners engage with the toolkit. There are also users that start playing, but get disengaged and do not follow the procedure. This leads to self-selected distribution between players and non-players (control), which presents the draw-back of non-random selection and the need to consider predisposition (willingness) to engage. To ensure real-world relevance, it is desirable that study measures are integrated with final class appraisal.

A third and final identified difference is that in the case of class studies, competition and discussion happen between people that know each other in the real world, just that they use the affordances of the software for communication.

Similarly to the way Swanson and Jhala (2012) conducted their studies, particicpants could be recruited through MTurk. In these controlled experiments it would be difficult to measure learning, however investigation into engagement is feasible, as this was carried out in Studies 4 and 5. This makes them ideal for usability studies of key features, so that learner's interpretation of what is learned is also examined.

Bibliography

Adams, A., Sasse, M.A., 1999. Users are not the enemy: Why users compromise security mechanisms and how to take remedial measures. Commun. ACM 42, 40–46. doi:10.1145/322796.322806

Åkerlind, G.S., 2005. Variation and commonality in phenomenographic research methods. Higher Education Research & Development 24, 321–334. doi:10.1080/07294360500284672

Albrechtsen, E., Hovden, J., 2009. The information security digital divide between information security managers and users. Computers & Security 28, 476–490. doi:10.1016/j.cose.2009.01.003

Aldrich, C., 2003. Simulations and the Future of Learning: An Innovative (and Perhaps Revolutionary) Approach to e-Learning. Pfeiffer.

Aldrich, C., 2009a. Virtual Worlds, Simulations, and Games for Education: A Unifying View. Innovate 5.

Aldrich, C., 2009b. The Complete Guide to Simulations and Serious Games: How the Most Valuable Content Will be Created in the Age Beyond Gutenberg to Google, illustrated. ed, Pfeiffer Essential Resources for Training and HR Professionals. Jossey Bass.

Andersen, B., Fradinho, M., Lefrere, P., Niitamo, V.-P., 2009. The Coming Revolution in Competence Development: Using Serious Games to Improve Cross-Cultural Skills, in: Online Communities and Social Computing. pp. 413–422.

Andersen, E., O'Rourke, E., Liu, Y.E., Snider, R., Lowdermilk, J., Truong, D., Cooper, S., Popovic, Z., 2012. The impact of tutorials on games of varying complexity, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12. ACM, Austin, Texas, USA, pp. 59–68. doi:10.1145/2207676.2207687

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., Wittrock, M.C., 2000. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition, 2nd ed. Pearson.

Anderson, P.H., Lawton, L., 1992. A Survey of Methods Used for Evaluating Student Performance on Business Simulations. Simulation & Gaming 23, 490–498. doi:10.1177/1046878192234009

Anderson, P.H., Lawton, L., 2009. Business Simulations and Cognitive Learning: Developments, Desires, and Future Directions. Simulation Gaming 40, 193–216. doi:10.1177/1046878108321624

Anderson, R., 2013. Liveblogging SOUPS 2013. Light Blue Touchpaper: Security Research, Computer Laboratory, University of Cambridge.

Anderson, T., Shattuck, J., 2012. Design-Based Research: A Decade of Progress in Education Research? Educational Researcher 41, 16–25. doi:10.3102/0013189x11428813

Aronson, J., 1994. A Pragmatic View of Thematic Analysis. The Qualitative Report 2.

Artino, A.R., 2008. Cognitive Load Theory and the Role of Learner Experience: An Abbreviated Review for Educational Practitioners. AACE Journal 16, 425–439.

Atherton, J.S., 2013a. Learning and Teaching; The Problem of Assessment [WWW Document]. URL http://www.learningandteaching.info/teaching/assess_problem.htm (accessed 12.4.13).

Atherton, J.S., 2013b. Learning and Teaching: Forms of Assessment [WWW Document]. URL http://www.learningandteaching.info/teaching/assess_form.htm (accessed 12.4.13).

Atherton, J.S., Hadfield, P., Meyers, R., 2008. Threshold Concepts in the Wild, in: Threshold Concepts: From Theory to Practice.

Attwell, G., 2007. Personal Learning Environments - the future of eLearning? eLearning Papers 2.

Aylett, R., 1999. Narrative in virtual environments-towards emergent narrative, in: Narrative Intelligence Symposium.

Aylett, R., Louchart, S., Tychsen, A., Hitchens, M., Figueiredo, R., Mata, C.D., 2007. Managing emergent character-based narrative, in: INTETAIN '08: Proceedings of the 2nd International Conference on INtelligent TEchnologies for Interactive enterTAINment. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium, pp. 1–8.

Baddeley, A., 1992. Working memory. Science 255, 556–559. doi:10.1126/science.1736359

Barab, S., Squire, K., 2004. Design-Based Research: Putting a Stake in the Ground. Journal of the Learning Sciences 13, 1–14. doi:10.1207/s15327809jls1301_1

Barnard-Wills, D., Ashenden, D., 2013. Playing with Privacy: Games for Education and Communication in the Politics of Online Privacy. Polit Stud n/a. doi:10.1111/1467-9248.12049

Barrow, D., Mitrovic, A., Ohlsson, S., Grimley, M., 2008. Assessing the Impact of Positive Feedback in Constraint-Based Tutors, in: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (Eds.), Intelligent Tutoring Systems,

Lecture Notes in Computer Science. Springer Berlin Heidelberg, Montreal, Canada, pp. 250–259.
Bartle, R.A., 1996. Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs. Journal of MUD Research 1.
Beautement, A., Sasse, M.A., Wonham, M., 2008. The compliance budget: managing security behaviour in organisations, in: Proceedings of the 2008 Workshop on New Security Paradigms, NSPW '08. ACM, Lake Tahoe, California, USA, pp. 47–58. doi:10.1145/1595676.1595684
Biggs, J.B., 1994. Modes of learning, forms of knowing, and ways of schooling, in: Andreas Demetriou, M.S. (Ed.), Neo-Piagetian Theories of Cognitive Development: Implications and Applications for Education. Routledge.
Biggs, J.B., 1995. Assessing for learning : some dimensions underlying new approaches to educational assessment. Alberta journal of educational research 41, 1–17.
Biggs, J.B., 2003. Aligning Teaching and Assessing to Course Objectives, in: Teaching and Learning in Higher Education: New Trends and Innovations.
Biggs, J.B., Collis, K.F., 1982. Evaluating the Quality of Learning: The Solo Taxonomy : Structure of the Observed Learning Outcome (Educational Psychology Series). Academic Pr.
Biggs, J.B., Tang, C., 2007. Teaching for Quality Learning at University (Society for Research Into Highter Education), 3rd ed. Open University Press.
Birks, D., Townsley, M., Stewart, A., 2012. Generative Explanations of Crime: Using Simulation to Test Criminological Theory. Criminology 50, 221–254. doi:10.1111/j.1745-9125.2011.00258.x
Blackwell, C., 2012. A Strategy for Formalizing Attack Patterns, in: The First International Workshop on Cyberpatterns Unifying Design Patterns with Security, Attack and Forensic Patterns.
Bloom, B.S., Englehart, M.D., Furst, E.J., Hill, W.H., Krathwohl, D., 1956. Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: The Cognitive Domain. Susan Fauer Company, Inc.
Bocconi, S., Bachvarova, Y., Ruskov, M., Fradinho Oliveira, M., 2012. Dealing with Threshold Concepts in Serious Games for Competence Development, in: Ma, M., Fradinho Oliveira, M., Baalsrud Hauge, J., Duin, H., Thoben, K.-D. (Eds.), Serious Games Development and Applications, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 158–169.
Boot, W.R., Blakely, D.P., Simons, D.J., 2011. Do Action Video Games Improve Perception and Cognition? Frontiers in Cognition 2.
Bransford, J.D., Schwartz, D.L., 1999. Rethinking Transfer: A Simple Proposal with Multiple Implications. Review of Research in Education 24, 61+. doi:10.2307/1167267
Bray, D.A., Croxson, K., Dutton, W.H., Konsynski, B., 2007. Sermo: A Community-Based, Knowledge Ecosystem. Social Science Research Network Working Paper Series. doi:10.2139/ssrn.1016483
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, 624–634. doi:10.1016/j.jesp.2009.02.016
Broudy, H.S., 1977. Types of knowledge and purposes of education, in: Anderson, R.C., Spiro, R.J., Montague, W.E. (Eds.), . Lawrence Erlbaum, pp. 1–17.
Brown, J.S., Collins, A., Duguid, P., 1989. Situated Cognition and the Culture of Learning. Educational Researcher 18, 32–42. doi:10.3102/0013189X018001032
Bullock, K., Tilley, N. (Eds.), 2011. Crime Reduction and Problem-oriented Policing (Crime Science Series), illustrated edition. ed. Willan.
Bull, S., Ahmad, N., Johnson, M., Johan, R., Mabbott, A., Kerly, A., 2008. Adaptive Navigation Support, Learner Control and Open Learner Models, in: Nejdl, W., Kay, J., Pu, P., Herder, E. (Eds.), Adaptive Hypermedia and Adaptive Web-Based Systems, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 275–278.
Bull, S., Mabbott, A., 2006. 20000 Inspections of a Domain-Independent Open Learner Model with Individual and Comparison Views, in: Ikeda, M., Ashley, K.D., Chan, T.-W. (Eds.), Intelligent Tutoring Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 422–432.
Butcher, P.G., Jordan, S.E., 2010. A comparison of human and computer marking of short free-text student responses. Computers & Education. doi:10.1016/j.compedu.2010.02.012
Cameron, J., Banko, K.M., Pierce, W.D., 2001. Pervasive negative effects of rewards on intrinsic motivation: The myth continues. The Behavior Analyst 24, 1–44.
Cappelli, D.M., Moore, A.P., Trzeciak, R.F., 2012. The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes (Theft, Sabotage, Fraud) (SEI Series in Software Engineering), 1st ed. Addison-Wesley Professional.
Chaiken, S., 1987. The Heuristic Model of Persuasion, in: Zanna, M.P. (Ed.), Personality and Social Psychology:

Ontario Symposium Proceedings. New York University, pp. 3–40.
Chandler, P., Sweller, J., 1991. Cognitive Load Theory and the Format of Instruction. Cognition and Instruction 8, 293–332. doi:10.2307/3233596
Cheng, M., Ho, C., 2008. A Study on Applying the Variation Theory to Chinese Communicative Writing. Asian Social Science 4, 14–29.
Clarke, R.V., Eck, J.E., 2009. Crime Analysis for Problem Solvers in 60 Small Steps. U.S. Department of Justice, Washington, D.C.
Clark, W., Gantt, H.L., 2010. The Gantt chart, a working tool of management; Nabu Press.
Cocea, M., 2006. Assessment of Motivation in Online Learning Environments, in: Adaptive Hypermedia and Adaptive Web-Based Systems. pp. 414–418.
Cocea, M., Magoulas, G.D., 2009. Hybrid model for learner modelling and feedback prioritisation in exploratory learning. Int. J. Hybrid Intell. Syst. 6, 211–230.
Cocea, M., Weibelzahl, S., 2009. Log file analysis for disengagement detection in e-Learning environments. User Modeling and User-Adapted Interaction 19, 341–385. doi:10.1007/s11257-009-9065-5
Collins, B.S., Mansell, R., 2004. Cyber trust and crime prevention: a synthesis of the state-of-the-art science reviews. London, UK.
Collins, H.M., 2001. Tacit Knowledge, Trust and the Q of Sapphire. Social Studies of Science 31, 71–85. doi:10.1177/030631201031001004
Conati, C., Gertner, A., Vanlehn, K., 2002. Using Bayesian Networks to Manage Uncertainty in Student Modeling. User Modeling and User-Adapted Interaction 12, 371–417. doi:10.1023/A:1021258506583
Cone, B., Irvine, C.E., Thompson, M.F., Nguyen, T., 2007. A video game for cyber security training and awareness. Computers & Security 26, 63–72. doi:10.1016/j.cose.2006.10.005
Cousin, G., 2008. Threshold Concepts: Old Wine in New Bottles or New Forms of Transactional Inquiry?, in: Land, R., Meyer, J.H.F., Smith, J. (Eds.), Threshold Concepts within the Disciplines. Sense Publishers, pp. 261–272.
Cox, A.L., Cairns, P., Berthouze, N., Jennett, C., 2006. The Use of Eyetracking for Measuring Immersion, in: Cognitive Science.
Crawford, C., 2004. Chris Crawford on Interactive Storytelling, illustrated edition. ed. New Riders Games.
Csikszentmihalyi, M., 1991. Flow: The Psychology of Optimal Experience. Harper Perennial.
Csikszentmihalyi, M., LeFevre, J., 1989. Optimal experience in work and leisure. Journal of personality and social psychology 56, 815–822.
Davies, P., 2003. Threshold Concepts: how can we recognise them?, in: Overcoming Barriers to Student Understanding: Threshold Concepts and Troublesome Knowledge.
Davies, P., Mangan, J., 2006. Embedding Threshold Concepts: from theory to pedagogical principles to learning activities, in: Threshold Concepts within the Disciplines Symposium. IEPR Staffordshire University.
De Bono, E., 1999. Six Thinking Hats, 2nd ed. Back Bay Books.
Deci, E.L., 1971. Effects of externally mediated rewards on intrinsic motivation. Journal of Personality and Social Psychology 18, 105–115. doi:10.1037/h0030644
Deci, E.L., Koestner, R., Ryan, R.M., 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological bulletin 125.
De Freitas, S., Jarvis, S., 2006. A Framework for Developing Serious Games to meet Learner Needs. The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) 2006.
De Freitas, S., Oliver, M., 2006. How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? Comput. Educ. 46, 249–264. doi:10.1016/j.compedu.2005.11.007
Denning, T., Kohno, T., Shostack, A., 2012. Control-Alt-Hack: a card game for computer security outreach, education, and fun. ACM, Denver, Colorado, USA.
Derry, S.J., 1996. Cognitive schema theory in the constructivist debate. Educational Psychologist 31, 163–174. doi:10.1080/00461520.1996.9653264
Dickey, M., 2005. Engaging by design: How engagement strategies in popular computer and video games can inform instructional design. Educational Technology Research and Development 53, 67–83. doi:10.1007/BF02504866
Dienes, Z.P., 1973. A theory of Mathematics Learning, in: Crosswhite, F.J., Higgins, J., Osborne, A.R., Shumway, R.J. (Eds.), Teaching Mathematics: Psychological Foundation. C. A. Jones Pub. Co.
Di Eugenio, B., Fossati, D., Yu, D., Haller, S., Glass, M., 2005. Aggregation Improves Learning: Experiments in Natural Language Generation for Intelligent Tutoring Systems, in: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL '05. Association for Computational Linguistics, Ann Arbor,

Michigan, pp. 50–57. doi:10.3115/1219840.1219847
Dunbar, N.E., Wilson, S.N., Adame, B.J., Elizondo, J., Jensen, M.L., Miller, C.H., Kauffman, A.A., Seltsam, T., Bessarabova, E., Vincent, C., Straub, S.K., Ralston, R., Dulawan, C.L., Ramirez, D., Squire, K., Valacich, J.S., Burgoon, J.K., 2013. MACBETH: Development of a Training Game for the Mitigation of Cognitive Bias. International Journal of Game-Based Learning 3. doi:10.4018/ijgbl.2013100102
Edwards, R., 2001. GNS and Other Matters of Role-playing Theory [WWW Document]. URL http://www.indie-rpgs.com/articles/1/
Egri, L., 2007. The Art of Dramatic Writing. Wildside Press.
Ekblom, P., 1997. Gearing Up Against Crime: A Dynamic Framework to Help Designers Keep Up With the Adaptive Criminal in a Changing World. International Journal of Risk, Security and Crime Prevention 2, 249–265.
Ekblom, P., 2001. The Conjunction of Criminal Opportunity: a Framework for Crime Reduction Toolkits. UK national Crime Reduction Website.
Ekblom, P., 2007. Making Offenders Richer, in: Farrell, G., Bowers, K., Johnson, S., Townsley, M. (Eds.), Imagination for Crime Prevention: Essays in Honour of Ken Pease. Criminal Justice Press, pp. 41–58.
Ekblom, P., 2011a. Crime Prevention, Security and Community Safety Using the 5Is Framework (Crime Prevention and Security Management). Palgrave Macmillan.
Ekblom, P., 2011b. Happy returns: ideas brought back from situational crime prevention's exploration of design against crime, in: Farrell, G., Tilley, N. (Eds.), The Reasoning Criminologist: Essays in Honour of Ronald V. Clarke, Crime Science. Routledge, p. 163+.
Elgood, C., 1997. Handbook of Management Games and Simulations, 6 Sub. ed. Gower Publishing Company.
Falchikov, N., Boud, D., 1989. Student Self-Assessment in Higher Education: A Meta-Analysis. REVIEW OF EDUCATIONAL RESEARCH 59, 395–430. doi:10.3102/00346543059004395
Fang, Y.-H., Chiu, C.-M., 2009. In justice we trust: Exploring knowledge-sharing continuance intentions in virtual communities of practice. Computers in Human Behavior.
Fanning, R.M., Gaba, D.M., 2007. The role of debriefing in simulation-based learning. Simulation in healthcare : journal of the Society for Simulation in Healthcare 2, 115–125. doi:10.1097/sih.0b013e3180315539
Fisher, R., Ury, W.L., 1991. Getting to Yes: Negotiating Agreement Without Giving In, 2nd Edition. ed. Penguin (Non-Classics).
Flanagan, M., Hokstad, L.M., Zimmermann, M., Ackermann, G., Fradinho, M., Andersen, B., 2010. Transformational learning and serious game design, in: Third Biennial Threshold Concepts Symposium: Exploring Transformative Dimensions of Threshold Concepts. Sydney, Australia.
Fogg, B.J., 2002. Persuasive Technology: Using Computers to Change What We Think and Do (Interactive Technologies), 1st ed. Morgan Kaufmann.
Fogg, B.J., 2009. A Behavior Model for Persuasive Design, in: Proceedings of the 4th International Conference on Persuasive Technology, Persuasive '09. ACM, Claremont, California, pp. 1–7. doi:10.1145/1541948.1541999
Forget, A., Chiasson, S., Biddle, R., 2007. Persuasion as Education for Computer Security, in: Bastiaens, T., Carliner, S. (Eds.), Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2007. AACE, Quebec City, Canada, pp. 822–829.
Fradinho, M.O., Andersen, B., Pereira, J., Seager, W., Ribeiro, C., 2011. The Use of Integrative Framework to Support the Development of Competences, in: Ma, M., Fradinho Oliveira, M., Madeiras Pereira, J. (Eds.), Serious Games Development and Applications, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 117–128.
Fraser, D., Linder, C., 2009. Teaching in higher education through the use of variation: examples from distillation, physics and process dynamics. European Journal of Engineering Education 34, 369–381. doi:10.1080/03043790902989507
Gagne, R.M., 1985. The Conditions of Learning and Theory of Instruction, 4 Sub. ed. Wadsworth Pub Co.
Gibson, D., 2007. simSchool and the Conceptual Assessment Framework, in: Gibson, D., Aldrich, C., Prensky, M. (Eds.), Games And Simulations in Online Learning: Research and Development Frameworks. IGI Global, pp. 308–322.
Gibson, D.E., 2003. Developing the Professional Self-Concept: Role Model Construals in Early, Middle, and Late Career Stages. Organization Science 14, 591–610. doi:10.2307/4135151
Gijbels, D., Dochy, F., Van den Bossche, P., Segers, M., 2005. Effects of Problem-Based Learning: A Meta-Analysis From the Angle of Assessment. Review of Educational Research 75, 27–61. doi:10.3102/00346543075001027
Greenberg, E., 1992. Creativity, Autonomy, and Evaluation of Creative Work: Artistic Workers In Organizations. The Journal of Creative Behavior 26, 75–80. doi:10.1002/j.2162-6057.1992.tb01162.x

Gunter, G.A., Kenny, R.F., Vick, E.H., 2006. A Case for a Formal Design Paradigm for Serious Games. International Digital Media and Arts Association 3.
Hansen, C., McCalla, G., 2003. Active Open Learner Modelling. AIED03.
Hartley, D., Mitrovic, A., 2002. Supporting Learning by Opening the Student Model, in: Intelligent Tutoring Systems : 6th International Conference, ITS 2002, Biarritz, France and San Sebastian, Spain, June 2-7, 2002. Proceedings. pp. 185–191.
Hartley, J., Betts, L.R., 2010. Four layouts and a finding: the effects of changes in the order of the verbal labels and numerical values on Likert-type scales. International Journal of Social Research Methodology 13, 17–27. doi:10.1080/13645570802648077
Hay, D.B., Kehoe, C., Miquel, M.E., Hatzipanagos, S., Kinchin, I.M., Keevil, S.F., Lygo-Baker, S., 2008. Measuring the quality of e-learning. British Journal of Educational Technology 39, 1037–1056. doi:10.1111/j.1467-8535.2007.00777.x
Hay, D., Kinchin, I., 2008. Using concept mapping to measure learning quality. Education + Training 167–182. doi:10.1108/00400910810862146
Hay, D., Kinchin, I., Lygo-Baker, S., 2008. Making learning visible: the role of concept mapping in higher education. Studies in Higher Education 295–311. doi:10.1080/03075070802049251
Hays, R.T., 2005. The Effectiveness of Instructional Games: A Literature Review and Discussion.
Herrington, J., McKenney, S., Reeves, T., Oliver, R., 2007. Design-based research and doctoral students: Guidelines for preparing a dissertation proposal. World Conference on Educational Multimedia, Hypermedia and Telecommunications 25–29.
Hmelo-Silver, C.E., Duncan, R.G., Chinn, C.A., 2007. Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). Educational Psychologist 42, 99–107.
Hoffler, T., Leutner, D., 2007. Instructional animation versus static pictures: A meta-analysis. Learning and Instruction 17, 722–738. doi:10.1016/j.learninstruc.2007.09.013
Hulpuş, I., Fradinho, M., Hayes, C., Hokstad, L., Seager, W., Flanagan, M., 2010. Rapid Competence Development in Serious Games: Using Case-Based Reasoning and Threshold Concepts, in: CSEDU.
Ijsselsteijn, W.A., de Kort, Y.A.W., Poels, K., in press. The Game Experience Questionnaire:Development of a self-report measure to assess the psychological impact of digital games. Manuscript in Preparation.
Irvine, C.E., Thompson, M.F., 2004. Expressing an Information Security Policy Within a Security Simulation Game.
Isbister, K., Flanagan, M., Hash, C., 2010. Designing Games for Learning: Insights from Conversations with Designers, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10. ACM, Atlanta, Georgia, USA, pp. 2041–2044. doi:10.1145/1753326.1753637
Iurgel, I., 2006. Cyranus - An Authoring Tool for Interactive Edutainment Applications, in: Technologies for E-Learning and Digital Entertainment, Lecture Notes in Computer Science. pp. 577–580.
Jones, J., Yuan, X., Carr, E., Yu, H., 2010. A comparative study of CyberCIEGE game and Department of Defense Information Assurance Awareness video, in: IEEE SoutheastCon 2010 (SoutheastCon), Proceedings of the. IEEE, Concord, NC, USA, pp. 176–180. doi:10.1109/secon.2010.5453895
Jurafsky, D., Martin, J.H., 2000. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition (Prentice Hall Series in Artificial Intelligence), 1st ed. Prentice Hall.
Juul, J., 2007. Variation over Time, in: von Borries, F., Walz, S.P., Böttger, M. (Eds.), Space Time Play. Birkhäuser.
Kanawattanachai, P., Yoo, Y., 2002. Dynamic nature of trust in virtual teams. The Journal of Strategic Information Systems 11, 187–213. doi:10.1016/S0963-8687(02)00019-7
Karpouzis, K., Yannakakis, G., Paiva, A., Nielsen, J.H., Vasalou, A., Jhala, A., 2013. User Modelling and Adaptive, Natural Interaction for Conflict Resolution, in: Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on. IEEE, pp. 719–721. doi:10.1109/acii.2013.131
Kebritchi, M., Hirumi, A., 2008. Examining the pedagogical foundations of modern educational computer games. Computers & Education 51, 1729–1743. doi:10.1016/j.compedu.2008.05.004
Keller, J.M., Kopp, T.W., 1987. An application of the ARCS Model of Motivational Design, in: Reigeluth, C.M. (Ed.), . Routledge, pp. 289–320.
Kendall, M.G., Gibbons, J.D., 1990. Rank Correlation Methods, 5th ed, Charles Griffin Book. Oxford University Press, USA.
Kickmeier-Rust, M.D., Albert, D., 2010. Micro-adaptivity: protecting immersion in didactically adaptive digital educational games. Journal of Computer Assisted Learning 26, 95–105. doi:10.1111/j.1365-2729.2009.00332.x

King, A., Staffieri, A., Adelgais, A., 1998. Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. Journal of Educational Psychology 90, 134–152.
Kirlappos, I., Beautement, A., Sasse, M.A., 2013. “Comply or Die” is dead: Long live security-aware principal agents, in: Workshop on Usable Security.
Kirschner, P.A., Sweller, J., Clark, R.E., 2006. Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. Educational Psychologist 41, 75–86. doi:10.1207/s15326985ep4102_1
Kniberg, H., 2007. Scrum and XP from the Trenches (Enterprise Software Development). Lulu.com.
Knutsson, J., Clarke, R.V. (Eds.), 2006. Putting Theory to Work: Implementing Situational Prevention and Problem-Oriented Policing. Criminal Justice Press.
Kollar, I., Fischer, F., Hesse, F., 2006. Collaboration Scripts – A Conceptual Analysis. Educational Psychology Review 18, 159–185. doi:10.1007/s10648-006-9007-2
Kompen, R.T., Edirisingha, P., Monguet, J.M., 2009. Using Web 2.0 Applications as Supporting Tools for Personal Learning Environments, in: Best Practices for the Knowledge Society. Knowledge, Learning, Development and Technology for All. pp. 33–40.
Kreitmayer, S., Rogers, Y., Laney, R., Peake, S., 2012. From Participatory to Contributory Simulations: Changing the Game in the Classroom, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’12. ACM, Austin, Texas, USA, pp. 49–58. doi:10.1145/2207676.2207685
Land, R., Meyer, J.H.F., Smith, J., 2008. Threshold Concepts within the Disciplines. Sense Publishers.
Laurillard, D., 2001. Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies, 2nd ed. Routledge.
Laurillard, D., 2009. The pedagogical challenges to collaborative technologies. International Journal of Computer-Supported Collaborative Learning 4, 5–20. doi:10.1007/s11412-008-9056-2
Lave, J., Wenger, E., 1991. Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives), 1st ed. Cambridge University Press.
Lawrence, J.S.S., Kirksey, W.A., Moore, T., 1983. External validity of role play assessment of assertive behavior. Journal of Psychopathology and Behavioral Assessment 5, 25–34. doi:10.1007/BF01343636
Leahy, W., Sweller, J., 2008. The imagination effect increases with an increased intrinsic cognitive load. Applied Cognitive Psychology 22, 273–283. doi:10.1002/acp.1373
Lee, J., 2008. A Hong Kong case of lesson study—Benefits and concerns. Teaching and Teacher Education 24, 1115–1124. doi:10.1016/j.tate.2007.10.007
Lindley, C.A., 2003. Game Taxonomies: A High Level Framework for Game Analysis and Design. Gamasutra.
Lo, M.L., Pong, W.Y., Chik, P.P., 2005. For Each And Everyone: Catering for Individual Differences Through Learning Studies, illustrated. ed. Hong Kong University Press.
Lucas, U., Mladenovic, R., 2007. The potential of threshold concepts: an emerging framework for educational research and practice. London Review of Education 5, 237–248. doi:10.1080/14748460701661294
Maier, F.H., Größler, A., 2000. What are we talking about? - A taxonomy of computer simulations to support learning. System Dynamics Review 16, 135–148. doi:10.1002/1099-1727(200022)16:2%3C135::aid-sdr193%3E3.0.co;2-p
Malheiros, M., Seager, W., Ruskov, M., Sasse, M.A., 2011. Individual and organizational perspectives on data use in serious games. University College London, London, UK.
Marsella, S.C., Johnson, L.W., Labore, C., 2000. Interactive pedagogical drama, in: AGENTS ’00: Proceedings of the Fourth International Conference on Autonomous Agents. ACM Press, New York, NY, USA, pp. 301–308.
Marton, F., 1981. Phenomenography — Describing conceptions of the world around us. Instructional Science 10, 177–200. doi:10.1007/bf00132516
Marton, F., 1992. Phenomenography and “the art of teaching all things to all men”. International Journal of Qualitative Studies in Education 5, 253–267. doi:10.1080/0951839920050305
Marton, F., Booth, S., 1997. Learning and Awareness (Educational Psychology Series). Routledge.
Marton, F., Pang, M.F., 2006. On Some Necessary Conditions of Learning. Journal of the Learning Sciences 15, 193–220. doi:10.1207/s15327809jls1502_2
Marton, F., Tsui, A.B.M., Chik, P.P.M., Ko, P.Y., Lo, M.L., 2004. Classroom Discourse and the Space of Learning. Routledge.
Mateas, M., 2001. A preliminary poetics for interactive drama and games. Digital Creativity 12, 140–152. doi:10.1076/digc.12.3.140.3224
Mateas, M., Stern, A., 2004. Natural Language Understanding in Façade: Surface-Text Processing, in: Technologies

for Interactive Digital Storytelling and Entertainment. pp. 3–13.
Mayer, R.E., 2004. Should There Be a Three-Strikes Rule Against Pure Discovery Learning? American Psychologist 59, 14–19.
McGonigal, J.E., Worthen, W.B., Niemeyer, G., Goldberg, K., Glazer, P., 2006. This Might Be a Game: Ubiquitous Play and Performance at the Turn of the Twenty-First Century. University of California, Berkeley.
McKenney, S., van den Akker, J., 2005. Computer-based support for curriculum designers: A case of developmental research. Educational Technology Research and Development 53, 41–66. doi:10.1007/bf02504865
Meyer, J.H.F., Land, R., 2003. Threshold Concepts and Troublesome Knowledge (1): linkages to ways of thinking and practising within the disciplines. Improving Student Learning – Ten Years On.
Meyer, J.H.F., Land, R., 2006a. Threshold concepts and troublesome knowledge: An introduction, in: Meyer, J.H.F., Land, R. (Eds.), Overcoming Barriers to Student Understanding: Threshold Concepts and Troublesome Knowledge. Routledge, pp. 3–18.
Meyer, J.H.F., Land, R., 2006b. Overcoming Barriers to Student Understanding: Threshold concepts and troublesome knowledge, 1st ed. Routledge.
Meyer, J.H.F., Land, R., 2006c. Threshold concepts and troublesome knowledge: issues of liminality, in: Meyer, J.H.F., Land, R. (Eds.), Overcoming Barriers to Student Understanding: Threshold Concepts and Troublesome Knowledge. Routledge, pp. 19–32.
Meyer, J.H.F., Land, R., Davies, P., 2008. Threshold concepts and troublesome knowledge (4): issues of variation and variability, in: Land, R., Meyer, J.H.F., Smith, J. (Eds.), Threshold Concepts within the Disciplines. Sense Publishers, pp. 59–74.
Moon, J.A., 2001. Reflection in Learning and Professional Development. Routledge.
Moreno, R., 2004. Decreasing Cognitive Load for Novice Students: Effects of Explanatory versus Corrective Feedback in Discovery-Based Multimedia. Instructional Science 32, 99–113–113. doi:10.1023/b:truc.0000021811.66966.1d
Murray, J.H., 1998. Hamlet on the Holodeck: The Future of Narrative in Cyberspace. The MIT Press.
Naaman, M., Boase, J., Lai, C.-H., 2010. Is it Really About Me? Message Content in Social Awareness Streams, in: Churchill, E. (Ed.), Computer Supported Collaborative Work. doi:10.1145/1718918.1718953
Nadolski, R.J., Hummel, H.G.K., van den Brink, H.J., Hoefakker, R.E., Slootmaker, A., Kurvers, H.J., Storm, J., 2008. EMERGO: A methodology and toolkit for developing serious games in higher education. Simulation & Gaming 39, 338–352. doi:10.1177/1046878108319278
Nakano, Y.I., Ishii, R., 2010. Estimating user's engagement from eye-gaze behaviors in human-agent conversations, in: Proceedings of the 14th International Conference on Intelligent User Interfaces, IUI '10. ACM, Hong Kong, China, pp. 139–148. doi:10.1145/1719970.1719990
Nicol, D., 2007. E-assessment by design: using multiple-choice tests to good effect. Journal of Further and Higher Education 31, 53–64. doi:10.1080/03098770601167922
Nonaka, I., Takeuchi, H., 1995. The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford University Press, USA.
Paas, F.G.W.C., van Merriënboer, J.J.G., 1994. Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills: A Cognitive-Load Approach. Journal of Educational Psychology 86, 122–133.
Pahnila, S., Siponen, M.T., Mahmood, A.A., 2007. Employees' Behavior towards IS Security Policy Compliance, in: System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on. IEEE, Waikoloa, HI, USA, p. 156b. doi:10.1109/hicss.2007.206
Palinko, O., Kun, A.L., Shyrokov, A., Heeman, P., 2010. Estimating cognitive load using remote eye tracking in a driving simulator, in: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, ETRA '10. ACM, Austin, Texas, pp. 141–144. doi:10.1145/1743666.1743701
Pang, M.F., 2003. Two Faces of Variation: on continuity in the phenomenographic movement. Scandinavian Journal of Educational Research 47, 145–156. doi:10.1080/00313830308612
Pang, M.F., Linder, C., Fraser, D., 2006. Beyond Lesson Studies and Design Experiments - Using Theoretical Tools in Practice and Finding Out How They Work. International Review of Economics Education 5, 28–45.
Pang, M.F., Marton, F., 2003. Beyond "lesson study": Comparing two ways of facilitating the grasp of some economic concepts. Instructional Science 31, 175–194. doi:10.1023/A:1023280619632
Pang, M.F., Marton, F., 2007. On the Paradox of Pedagogy: the relative contribution of teachers and learners to learning. Iskolakultúra Online 2007, 1–29.
Pang, M.F., Meyer, J.H.F., 2010. Modes of variation in pupils' apprehension of a threshold concept in Economics, in: Meyer, J.H.F., Land, R., Baillie, C. (Eds.), Threshold Concepts and Transformational Learning. Sense

Publishers, Rotterdam, Netherlands, pp. 365–381.
Paris, S.G., Lipson, M.Y., Wixson, K.K., 1983. Becoming a strategic reader. Contemporary Educational Psychology 8, 293–316. doi:10.1016/0361-476x(83)90018-8
Parker, D.B., 1998. Fighting Computer Crime: A New Framework for Protecting Information. Wiley.
Pea, R.D., 2004. The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity. Journal of the Learning Sciences 13, 423–451. doi:10.1207/s15327809jls1303_6
Perkins, D., 2006. Constructivism and troublesome knowledge, in: Meyer, J.H.F., Land, R. (Eds.), Overcoming Barriers to Student Understanding: Threshold Concepts and Troublesome Knowledge. Routledge.
Perkins, D., 2008. Beyond Understanding, in: Land, R., Meyer, J.H.F., Smith, J. (Eds.), . Sense Publishers.
Petty, R.E., Tormala, Z.L., Hawkins, C., Wegener, D.T., 2001. Motivation to Think and Order Effects in Persuasion: The Moderating Role of Chunking. Pers Soc Psychol Bull 27, 332–344. doi:10.1177/0146167201273007
Pink, D.H., 2011. Drive: The Surprising Truth about What Motivates Us. Canongate Books Ltd.
Ploegh, K., Tillema, H.H., Segers, M.S., 2009. In search of quality criteria in peer assessment practices. Studies In Educational Evaluation 35. doi:10.1016/j.stueduc.2009.05.001
Powers, D., 2002. Stumping e-rater:challenging the validity of automated essay scoring. Computers in Human Behavior 18, 103–134. doi:10.1016/S0747-5632(01)00052-8
Prensky, M., 2003. Digital game-based learning. Comput. Entertain. 1, 21. doi:10.1145/950566.950596
Price, M., Carroll, J., O'Donovan, B., Rust, C., 2010. If I was going there I wouldn't start from here: a critical commentary on current assessment practice. Assessment & Evaluation in Higher Education 36, 479–492. doi:10.1080/02602930903512883
Quilici, J.L., Mayer, R.E., 1996. Role of Examples in How Students Learn to Categorize Statistics Word Problems. Journal of Educational Psychology 88, 144–161.
Ranzijn, F.J.A., 1991. The Number of Video Examples and the Dispersion of Examples as Instructional Design Variables in Teaching Concepts. The Journal of Experimental Education 59, 320–330. doi:10.2307/20152300
Reason, P., Bradbury-Huan, H., 2001. Handbook of Action Research: Participative Inquiry and Practice, 1st ed. SAGE Publications Ltd.
Reeves, B., Malone, T., 2007. Leadership in Games and at Work: Implications for the Enterprise of Massively Multiplayer Online Role-playing Games [WWW Document]. URL http://www.seriosity.com/downloads/Leadership_In_Games_Seriosity_and_IBM.pdf
Reeves, T.C., 1999. A Research Agenda for Interactive Learning in the New Millennium, in: Collis, B., Oliver, R. (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 1999. AACE, Chesapeake, VA, pp. 15–20.
Reeves, T.C., 2006. Design research from a technology perspective. Educational Design Research 52–66.
Riedl, M.O., Stern, A., Dini, D., 2006. Mixing story and simulation in interactive narrative, in: In Proceedings of the Second Annual Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-06.
Riegelsberger, J., Sasse, M., McCarthy, J., 2005. The mechanics of trust: A framework for research and design. International Journal of Human-Computer Studies 62, 381–422. doi:10.1016/j.ijhcs.2005.01.001
Ruskov, M., 2005. Dialogue management systems for interactive storytelling.
Ruskov, M., Celdran, J.M., Ekblom, P., Sasse, M.A., 2012. Unlocking the Next Level of Crime Prevention: Development of a Game Prototype to Teach the Conjunction of Criminal Opportunity. Information Technologies and Control 10, 15–21.
Ruskov, M., Ekblom, P., Sasse, M.A., 2013. In Search for the Right Measure: Assessing Types of Developed Knowledge while Using a Gamified Web Toolkit, in: 7th European Conference on Games Based Learning.
Ruskov, M., Ekblom, P., Sasse, M.A., Bayley, I., Blackwell, C., 2014. Towards a Simulation of Information Security Behaviour in Organisations. Springer, Abingdon, Oxfordshire , UK.
Ruskov, M., Ruskov, P., 2006. Negotiation Design Pattern for Serious Games, in: 10th International Workshop on Experimental Interactive Learning in Industrial Management.
Ruskov, M., Seager, W., Sasse, M.A., 2010. Persuading Giants to be Wise: An Exploratory Study of Advice Sharing in Online Games. Presented at the The PLE Conference 2010.
Ruttkay, Z., Pelachaud, C. (Eds.), 2005. From Brows to Trust: Evaluating Embodied Conversational Agents, 1st ed, Human-Computer Interaction Series. Springer.
Ryan, M., 2008. 16 ways to use Second Life in your classroom: pedagogical approaches and virtual assignments.
Ryan, R., Rigby, C., Przybylski, A., 2006. The Motivational Pull of Video Games: A Self-Determination Theory Approach. Motivation and Emotion 30, 344–360. doi:10.1007/s11031-006-9051-8

Sauvé, L., Renaud, L., Kaufman, D., Marquis, J.-S., 2007. Distinguishing between games and simulations: A systematic review. Educational Technology & Society 10, 247–256.
Schilling, M.A., Vidal, P., Ployhart, R.E., Marangoni, A., 2003. Learning by Doing Something Else: Variation, Relatedness, and the Learning Curve. MANAGEMENT SCIENCE 49, 39–56. doi:10.1287/mnsc.49.1.39.12750
Schmidt, H.G., Loyens, S.M.M., Van Gog, T., Paas, F., 2007. Problem-Based Learning is Compatible with Human Cognitive Architecture: Commentary on Kirschner, Sweller, and Clark (2006). Educational Psychologist 42, 91–97. doi:10.1080/00461520701263350
Schmidt, R.A., 1975. A schema theory of discrete motor skill learning. Psychological Review 82, 225–260.
Schön, D.A., 1991. The Reflective Practitioner: How Professionals Think in Action (Arena), New edition. ed. Ashgate Publishing.
Schultz, 2002. A framework for understanding and predicting insider attacks. Computers & Security 21, 526–531. doi:10.1016/s0167-4048(02)01009-x
Schwartz, B., Ward, A., 2004. Doing Better but Feeling Worse: The Paradox of Choice. Positive Psychology in Practice 86–104. doi:10.1002/9780470939338.ch6
Scouller, K., 1998. The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. Higher Education 35, 453–472. doi:10.1023/a%253a1003196224280
Seager, W., Ruskov, M., Fradinho, M., Sasse, A., 2010. Eliciting and Modelling Expertise for Serious Game Design, in: Taisch, M., Cassina, J., Smeds, R. (Eds.), Experimental Learning on Sustainable Management, Economics and Industrial Engineering: Proceedings of 14th Workshop of the SIG on Experimental Interactive Learning in Industrial Management of the IFIP WG 5.7. Poliscript.
Searle, J.R., 1996. The Construction of Social Reality. Penguin Books Ltd.
Shafer, J., 2012. Lowering the Gates. Jon Shafer on Design.
Shermis, M.D., Burstein, J.C. (Eds.), 2002. Automated Essay Scoring: A Cross-disciplinary Perspective, 1st ed. Routledge.
Shostack, A., 2012. Elevation of Privilege: Drawing Developers into Threat Modeling.
Shuval, J., Adler, I., 1980. The role of models in professional socialization. Social Science & Medicine. Part A: Medical Psychology & Medical Sociology 14, 5–14. doi:10.1016/S0271-7123(80)90642-2
Sidor, S.M., 2008. Practiceware Works: Leadership Programs Without Comprehensive Practice Component Wastes Organizations' Time And Money - New Options for Training Functions Focused on Results.
Siegle, G.J., Ichikawa, N., Steinhauer, S., 2008. Blink before and after you think: blinks occur prior to and following cognitive load indexed by pupillary responses. Psychophysiology 45, 679–687. doi:10.1111/j.1469-8986.2008.00681.x
Si, M., Marsella, S.C., Pynadath, D.V., 2009. Directorial Control in a Decision-Theoretic Framework for Interactive Narrative, in: ICIDS '09: Proceedings of the 2nd Joint International Conference on Interactive Digital Storytelling. Springer-Verlag, Guimar\, aes, Portugal, pp. 221–233. doi:10.1007/978-3-642-10643-9_27
Slater, M., 2004. How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments. Presence: Teleoperators and Virtual Environments 13, 484–493. doi:10.1162/1054746041944849
Stajano, F., Wilson, P., 2011. Understanding scam victims: seven principles for systems security. Commun. ACM 54, 70–75. doi:10.1145/1897852.1897872
Standifer, R.L., Thiault, P., Pin, R., 2010. Leadership Development in an Electronic Frontier: Connecting Theory to Experiential Software Through Supplemental Materials. Journal of Leadership & Organizational Studies 17, 167–176. doi:10.1177/1548051810366714
Stash, N.V., Cristea, A.I., De Bra, P.M., 2004. Authoring of learning styles in adaptive hypermedia: problems and solutions, in: WWW Alt. '04: Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters. ACM, New York, NY, USA, pp. 114–123. doi:10.1145/1013367.1013387
Sugrue, B., 1995. A Theory-Based Framework for Assessing Domain-Specific Problem-Solving Ability. Educational Measurement: Issues and Practice 14, 29–35. doi:10.1111/j.1745-3992.1995.tb00865.x
Susi, T., Johannesson, M., Backlund, P., 2007. Serious Games: An Overview. University of Skovde, Sweden.
Swanson, R., Jhala, A., 2012. A Crowd-Sourced Collection of Narratives for Studying Conflict, in: Language, Resources and Evaluation Workshop on Computational Models of Narrative. Istanbul, Turkey.
Sweller, J., 2006. The worked example effect and human cognition. Learning and Instruction 16, 165–169. doi:10.1016/j.learninstruc.2006.02.005
Sweller, J., Kirschner, P.A., Clark, R.E., 2007. Why Minimally Guided Teaching Techniques Do Not Work: A Reply to Commentaries. Educational Psychologist 42, 115–121. doi:10.1080/00461520701263426
Tanenbaum, J.G., Bizzocchi, J., 2009. Close Reading Oblivion: Character Believability and Intelligent Personalization

in Games. Loading...: The Journal of the Canadian Game Studies Association 3.
Tanimoto, S., 2005. Dimensions of Transparency in Open Learner Models, in: Proc. Int'l Workshop on Learner Modelling for Reflection, to Support Learner Control, Metacognition and Improved Communic. between Teachers and Learners.
Tilley, N., 1993. After Kirkholt: Theory, method, and results of replication evaluations, Crime Prevention Unit Series. Home Office Police Dept.
Trinder, K., Guiller, J., Margaryan, A., Littlejohn, A., Nicol, D., 2008. Learning from digital natives: bridging formal and informal learning. The Higher Education Academy.
Tuglular, T., Spafford, E.H., 1997. A framework for characterisation of insider computer misuse.
Tychsen, A., Newman, K., Brolund, T., Hitchens, M., 2007. Cross-format analysis of the gaming experience in multi-player role-playing games, in: Proceedings of DiGRA.
Tychsen, A., Tosca, S., Brolund, T., 2006. Personalizing the Player Experience in MMORPGs, in: Göbel, S., Malkewitz, R., Iurgel, I. (Eds.), Technologies for Interactive Digital Storytelling and Entertainment. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 253–264.
Ulicsak, M., Wright, M., 2010. Games in Education: Serious Games. Futurelab.
Van den Akker, J., 1999. Principles and Methods of Development Research, in: van den Akker, J., Branch, R., Gustafson, K., Nieveen, N., Plomp, T. (Eds.), Design Approaches and Tools in Education and Training. Springer Netherlands, pp. 1–14.
Van den Hoogen, W., IJsselsteijn, W., de Kort, Y., 2008. Exploring Behavioral Expressions of Player Experience in Digital Games. Proceedings of the Facial and Bodily Expressions for Control and Adaptation of Games Workshop.
Van Merriënboer, J.J.G., Sweller, J., 2005. Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. Educational Psychology Review 17, 147–177. doi:10.1007/s10648-005-3951-0
Vasalou, A., Ingram, G., Khaled, R., 2012. User-centered Research in the Early Stages of a Learning Game, in: Proceedings of the Designing Interactive Systems Conference, DIS '12. ACM, Newcastle Upon Tyne, United Kingdom, pp. 116–125. doi:10.1145/2317956.2317976
Walker, M., 2009. An investigation into written comments on assignments: do students find them usable? Assessment & Evaluation in Higher Education 34, 67–78. doi:10.1080/02602930801895752
Weber, R.P., 1990. Basic Content Analysis (Quantitative Applications in the Social Sciences), 2 Sub. ed. Sage Publications, Inc.
Wenger, E., McDermott, R., Snyder, W.M., 2002. Cultivating Communities of Practice, 1st ed. Harvard Business Press.
Williamson, D.M., Mislevy, R.J., Bejar, I.I., 2006. Automated Scoring of Complex Tasks in Computer Based Testing, 1st ed. Lawrence Erlbaum Associates.
Wills, S., Leigh, E., Ip, A., 2010. The Power of Role-based e-Learning: Designing and Moderating Online Role Play (Connecting with E-learning), 1st ed. Routledge.
Wilson, S., Liber, O., Beauvoir, P., MIlligan, C., Johnson, M., Sharples, P., 2006. Personal Learning Environments: Challenging the dominant design of educational systems, in: 1. TENC: Publications and Preprints.
Winn, B.M., 2008. The Design, Play, and Experience Framework, in: Ferdig, R.E. (Ed.), Handbook of Research on Effective Electronic Gaming in Education. Michigan State University.
Wood, W., 1982. Retrieval of attitude-relevant information from memory: Effects on susceptibility to persuasion and on intrinsic motivation. Journal of Personality and Social Psychology 42, 798–810.
Wright, W., 2009. Motivation is more important than education [WWW Document]. Education Futures. URL http://www.educationfutures.com/2009/07/16/will-wright-motivation-is-more-important-than-education/ (accessed 3.25.10).
Xu, W., Rudnicky, A.I., 2000. Task-based dialog management using an agenda, in: ANLP/NAACL 2000 Workshop on Conversational Systems. Association for Computational Linguistics, Morristown, NJ, USA, pp. 42–47. doi:10.3115/1117562.1117571
Yee, N., 2005. Motivations of Play in MMORPGs, in: DiGRA 2005 Conference, Vancouver. Online Im Internet: Http://www. Gamesconference. Org.
Yee, N., Ducheneaut, N., Nelson, L., 2012. Online Gaming Motivations Scale: Development and Validation, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12. ACM, Austin, Texas, USA, pp. 2803–2806. doi:10.1145/2207676.2208681
Yukl, G.A., 2005. Leadership in Organizations (6th Edition), 6th ed. Prentice Hall.
Yusoff, A., Crowder, R., Gilbert, L., Wills, G., 2009. A Conceptual Framework for Serious Games, in: Advanced Learning Technologies, 2009. ICALT 2009. Ninth IEEE International Conference on. IEEE, pp. 21–23.

doi:10.1109/icalt.2009.19
Zeek, C., Foote, M., Walker, C., 2001. Teacher Stories and Transactional Inquiry: Hearing the Voices of Mentor Teachers. Journal of Teacher Education 52, 377–385. doi:10.1177/0022487101052005004
Zuckerman, M., Porac, J., Lathin, D., Deci, E.L., 1978. On the Importance of Self-Determination for Intrinsically-Motivated Behavior. Pers Soc Psychol Bull 4, 443–446. doi:10.1177/014616727800400317

Glossary

absorption – along with *flow*, *presence* and *immersion* psychological construct representing forms of engagement. These constructs are compared by (Brockmyer et al., 2009)

achievement – the drive many people have to achieve things, be it as a form of *advancement*, *discovery* or *competition*. Along with and *immersion* and *social motivation*, this is one of the three main groups of motivational drives to play massive multiplayer online games according to (Yee, 2005)

action research – research attempting to solve an immediate problem and reflecting on the process. Further explanation and references can be found in (Reason and Bradbury-Huan, 2001)

active identification – the process of people identifying with the behaviours of others that they see. This is one of three behaviours along with *active rejection* and *inactive observation* that have been observed when studying role-modeling behaviour in organisations Shuval and Adler (1980)

active rejection – the process of refusing to identify with observed behaviours of others. This is one of three behaviours along with *active identification* and *inactive observation* that have been observed when studying role-modeling behaviour in organisations Shuval and Adler (1980)

advert-games – alongside *learning games* a type of *serious games*, made for the purposes of advertising

agile development – a range of software development practices, notably *Scrum* and extreme programming, utilising iterative planning and participative management. Further explanation can be found in (Kniberg, 2007)

anticipation of risk, effort and reward – the rational assessment of the crime situation as made by the offender. One of the offender-oriented causes of crime, according to CCO (Ekblom, 2011a)

art games – alongside *learning games* a type of serious games, made as a form of interactive digital art

assessment fatigue – loss of rigour in responses to assessment questions as an effect of prolonged examination

attacker – in information security the person or organisation performing an attack.

autonomy – the concept representing freedom of choice, which, according to self-determination theory has positive effects on intrinsic motivation. Further explanation and references can be found in (Zuckerman et al., 1978)

backstory – see *story world*

bag-of-words – a text-analytical technique that ignores the structure of texts, and represents them as an unordered collections of the words that they contain, thus making such a representation equivalent to an integer vector space with each potential word featured as a dimension.

between-subjects design – along with *within-subjects design* one of the basic standard techniques to conduct experimental studies. In between-subjects design, participants are split in two or more groups and, everything else the same, are subjected to different interventions. This is done as a way to ensure that any observed differences between the groups (*dependent variable*) are an effect of differences between groups (*independent variable*). Due to the temporal order, it is commonly assumed that there is a causation from the independent to the dependent variable.

bibliographical knowledge – see *declarative knowledge*

bounded knowledge – knowledge, that has implications over a limited domain, and other knowledge is needed to explain beyond these limits. In the context of *threshold concepts* this leads to the idea that knowledge should be considered as *provisional*, thus always subject to further refinement. Further explanation and references can be found in (Meyer and Land, 2006a)

captology or **persuasive technologies** – the study and development of technologies intended to persuade someone to change a behaviour or attitude. It is introduced in (Fogg, 2002)

carrot-and-stick approach – a behavioural motivational approach that combines reinforcement techniques for what needs to be encouraged (carrot – positive) and negative penalisation (stick – negative) for what needs to be discouraged.

closed exam questions – see *multiple-choice questions*

coding or **coder** – the process or the person performing coding of qualitative data, whereas generally this could be interpreted as coding of a software programme, this second meaning is not used here

cognitive apprenticeship, also referred to as **shadowing** – learning from someone by helping them (Brown et al.,

1989)

cognitive load theory – a research theory about the way people learn. It is developed around the idea that people can process only a limited amount of information in their *working memory*, thus learning has to manage this limited capacity. A cornerstone finding is that cognitive load can be classified into *germane cognitive load* and *extraneous cognitive load*. Further explanation and reference can be found in (van Merriënboer and Sweller, 2005)

cognitive scaffolding, also referred to as **guidance during learning** – the process of providing learning support. It is now widely established that this increases the potential of the learner. Further explanation and references can be found in (Kirschner et al., 2006)

computer-supported collaborative learning – a sub-discipline of *human-computer interaction* that focuses on group learning with the help of computers

conditional knowledge, also referred to as **situational**, **proactive** or **contextual knowledge** – a form of knowledge that is traditionally less widely discussed, but is of key importance in the context of *situated cognition*. It has to do with the ability to decide if certain *declarative* or *procedural knowledge* can be applied in the specific situation. It could be explained as knowledge that tells know-when and know-where. It has been introduced by (Paris et al., 1983)

constructivism – the *epistemology* that postulates that knowledge is created by people. Also, a theory about learning that posits that knowledge is being constructed individually by learners, based on their surrounding environment. With the help of *cognitive scaffolding*, the learner could be allowed a wider potential for learning. This holds both for the invention of new theories and for the development of an understanding of already existing cognitive constructs

content analysis – a qualitative research method that focuses on the form of the studied data. Further explanation can be found in (Weber, 1990).

contextual knowledge – see *conditional knowledge*

contrast – along with *separation*, *generalisation* and *fusion*, one of four necessary conditions of learning according to (Marton and Pang, 2006).

conversational framework – a theoretical framework representing the learning process and its complexities originating from inefficiencies of communication and differences in perceptions. It has been introduced by Laurrilard (2001)

crime preventer – people that, with their presence, actions or attitude, make a crime less likely. One of the situational factors of crime, according to CCO (Ekblom, 2011a)

crime prevention – a subfield of *crime sciences* that deals with preventing crime

crime promoter – people that, with their presence, actions or attitude, make a crime more likely. One of the situational causes of crime, according to CCO (Ekblom, 2011a)

crime sciences – a recently emerged science of criminality that for the first time considers situational contributions towards crime

crime analysis triangle – a simple and widely used analytical tool in crime sciences. It has been introduced by (Clarke and Eck, 2009)

criminality (predisposition) – the inherent (or developed in early age) criminal predisposition of an *offender*. One of the offender-oriented causes of crime, according to CCO (Ekblom, 2011a)

data triangulation – evidence from different sources that are mutually reinforcing (Lucas and Mladenovic, 2007)

declarative knowledge, also referred to as **factual, descriptive, propositional**, **possessive**, **bibliographical knowledge** or **know-what knowledge** – a form of knowledge, whose existence is widely agreed on within all major epistemological communities. It is related to the memorisation and comprehension of facts and their descriptions. It is often explained as know-what as opposed to know-how. Further explanation and references can be found in (Gijbels et al., 2005)

deep learning – an approach to learning where a deeper understanding of the phenomenon is sought, trying to understand how it interacts with its environment. Contrast this to *surface learning*, where learners focus on memorising what was directly observed. Further explanation and references can be found in (Atherton 2013)

descriptive knowledge – see *declarative knowledge*

design-based research, also called **development research** or **developmental research** – a research approach focusing on improved development of learning interventions, rather than conducting scientific experiments (Anderson and Shattuck, 2012)

digital storytelling – see *interactive storytelling*

discourse analysis – analysis of qualitative (typically textual) data by means of exploring the communication between participants in some sort of interchange (e.g. conversation)

educational games – see *learning games*

emergent narrative – an approach to *interactive storytelling*, that does not rely on a predetermined dramatic arc. It is inspired by practice in role-playing games and is driven by users. It has been introduced by (Aylett, 1999)

enacted object of learning – the actual learning activities as they happen and learner's perceptions form them. It has been introduced by (Marton et al., 2004)

epistemology – the theory of the nature of knowedge and how it is being accumulated. Notably here the epistemology of *constructivism* is considered

essay writing – a reflection, *knowledge sharing* and assessment technique that involves writing a stylistically coherent text of up to several pages

experiential learning – the process of learning through experience. Examples of such learning are *problem-based learning*, *inquiry learning* or *game-based learning*

extended abstract answer – a class of answers relating several parallel themes and deriving conclusions about the described phenomenon according to the *SOLO hierarchy*. Such answers are the most advanced in the hierarchy. It has been introduced by (Biggs and Tang, 2007)

externalisation – see *knowledge externalisation*

extraneous cognitive load – according to *cognitive load theory* one of the two principle types of cognitive load along with *germane cognitive load*. Represents the cognitive load that is not crucial to the learning activity and thus should be reduced. Further explanation and reference can be found in (van Merriënboer and Sweller, 2005)

factual knowledge – see *declarative knowledge*

flow – a psychological state of mind that is associated with extreme involvement, a sense of being in control and experiencing time distortions. Flow is achieved through a balance between perceived skills and perceived challenges. It has been introduced by (Csikszentmihalyi, 1991) and compared with other psychological constructs representing engagement by (Brockmyer et al., 2009)

formative assessment – assessment within a learning process that is being conducted before its end in order to provide corrective feedback. Further explanation and references can be found in (Atherton, 2013b)

fusion – the experience of the full range of possible interactions between different aspects of the learned phenomenon. Along with *contrast*, *separation* and *generalisation*, one of four necessary conditions of learning according to (Marton and Pang, 2006)

game-based learning, commonly abbreviated as GBL – learning with the help of games. Does not specify how exactly learning with games happens, thus includes both learning while creating and learning while playing games

game design – the craft of developing engaging games, typically the focus is on player interactions

game design triangle – an empirical model of three game design components: *ludology* (play), *simulation* and narrative (storytelling). Further explanation can be found e.g. in (Edwards, 2001)

game mechanics – the clockworks of a game. Involves a formal programmable model, with complicated and percievably not interesting parts often hidden from the user

gamification – the idea of quantifying progress as a way of motivating people Results in trackable (and thus more tangible) progress, possibility to compare and compete with others

Gantt chart – a diagramming technique used in project management to represent the temporal interdependence of tasks within the project. It is named after its author and is described in (Clark and Gantt, 2010)

generalisation – Along with *contrast*, *separation* and *fusion*, one of the four techniques to experience aspects of learning variation. Involves fixating a feature of variation to a certain value and letting its surrounding features vary to see this value in different contexts. It has been introduced by (Marton and Pang, 2006)

germane cognitive load – according to *cognitive load theory* one of the two principle types of cognitive load along

with *extraneous cognitive load*. Represents the cognitive load that is key to the learning activity and thus should be facilitated. Further explanation and reference can be found in (van Merriënboer and Sweller, 2005)

human-computer interaction, commonly abbreviated as *HCI* – the discipline within computer science that deals with interaction with computers, technology usability, and user experiences with technology

illocutionary force, also referred to as **conversational pragmatics** – the role a statement or utterance has in the context of a conversation. Seen as a complement to syntax and semantics in a conversation (Searle, 1996)

immersion – along with *absorption*, *flow* and *presence* psychological construct representing engagement. These constructs are compared by (Brockmyer et al., 2009). Also, along with and *achievement* and *social motivation*, the willingness to experience immersion is one of the three main groups of motivational drives to play massive multiplayer online games according to (Yee, 2005)

imperative knowledge – see *procedural knowledge*

implementation – implementation of intervention, which in the context of crime prevention and CCO in particular could be a *security implementation* or in the case of *design-based research* could be a *software implementation*, but does not have to be

inactive observation – the process of unconscious perception that people perform while registering behaviours of others. This is one of three behaviours along with *active rejection* and *inactive observation* that have been observed when studying role-modeling behaviour in organisations Shuval and Adler (1980)

information chunking – the effect of subdividing information in smaller labelled pieces. Further explanation and references can be found in (Petty et al., 2001)

information security or **cybersecurity** – the craft (and more recently science) of protecting information that has developed in the context of internet

inquiry learning, also referred to as **inquiry-based learning** – a pedagogical principle that is centred around learner's enquiries, as opposed to a predefined learning plan. Further explanation and references can be found in (Hmelo-Silver et al., 2007)

insider, or **insider attacker** – in information security a person with legitimate access to the system, who performs an *attack* from inside the system

instructional games – see *learning games*

integrative knowledge – a piece of knowledge that once grasped, helps learners to understand interrelationships that before that were hidden to them. This is one of the five typical characteristics of *threshold concepts*. Further explanation and references can be found in (Meyer and Land, 2006a)

intelligent tutoring systems, often abbreviated as *ITS* – computer systems developed to serve as tutors to learners. Often they involve a *virtual character* that imitates a human tutor

intended object of learning – the planned learning activities that would support students. It has been introduced by (Marton et al., 2004)

interactive and digital storytelling, also referred to as either **interactive** or **digital storytelling** – computer-based storytelling applications that usually involve the learner in the creation of a story. It also typically involves *virtual characters*. This is further discussed in (Crawford, 2004)

interest – the implicit goals of one side in a negotiation. This has been contrasted to *position*, which represents the declared goals. These two are discussed in (Fisher and Ury, 1991)

invariance – in variation theory the features of the explored object that are kept constant for the sake of keeping them away from the learner's attention. Phenomenographic studies have shown that invariance results in learners taking these features for granted or not noticing them

irreversible knowledge – a piece of knowledge that, once grasped, is difficult to disregard. This is one of the five typical characteristics of *threshold concepts*. Further explanation and references can be found in (Meyer and Land, 2006a)

knowledge externalisation – the process of turning tacit knowledge into explicit (Nonaka and Takeuchi, 1995)

know-how – see *procedural knowedge*, discussed in (Biggs, 1994)

know-with – one of the proposed forms of conditional knowledge (Broudy, 1977)

know-what – see *declarative knowledge*

know-what-works – one of the proposed forms of conditional knowledge (Ekblom, 2011a, p. 29)

know-when-to-act – one of the proposed forms of conditional knowledge (Ekblom, 2011a, p. 29)

know-where – one of the proposed forms of conditional knowledge (Ekblom, 2011a, p. 29)

know-who-to-involve – one of the proposed forms of conditional knowledge (Ekblom, 2011a, p. 29)

know-why – one of the proposed forms of conditional knowledge (Ekblom, 2011a, p. 29)

learning by doing – the approach to learning through actual execution of the work, usually refers to improvement of work. It is a very good match with the theory of *situated cognition*. (Brown et al., 1989)

learning games, also referred to as **instructional, educational games** or **serious games for learning** or even confusingly simply *serious games* – games that have been designed for the purpose of learning. Along with *advert-games* and *art games*, are types of *serious games*. Further explanation and references can be found in (Kebritchi and Hirumi, 2008)

learning study – a *phenomenographic* approach developed from the *lesson study* but focusing not on improvement only of the lesson, but also of the understanding of how students percieve what is learned. This is a type of *transactional inquiry* and was introduced in (Pang and Marton, 2003)

lesson study – a Japanese traditional method of lesson development, where a group of teachers attend each-other's classes and discuss how the presentation of the learning material could be improved. This is a type of *transactional inquiry*. Further explanation and references can be found in (Pang and Marton, 2003)

lived object of learning – the execution of planned learning activities that are intended to support students. It is one of the articulations of the *object of learning* and has been introduced by (Marton et al., 2004)

log-driven performance assessment – performance assessment based on logs (recorded traces) left in a computer system

ludology – concerned with the playful aspects of the game and thus complements simulation and narrative in the *game design triangle*. It is often related to competition. Further explanation can be found e.g. in (Edwards, 2001)

mechanics, or **mechanism** – could refer to *game mechanics* or *social (crime) mechanisms*

microadaptivity – an adaptive technique to unobtrusively assess learning and introduce real-time learning support. It has been introduced by (Kickmeier-Rust and Albert, 2010)

mimicry – an extreme form of *surface learning* when a learner is able to reproduce the form, but not the meaning. In cases when the form is sufficient representation of a piece of knowledge (e.g. MCQs), mimicry is indistinguishable. Further explanation can be found e.g. in (Atherton, 2013b)

multiple-choice questions, commonly abbreviated as *MCQ*, and also referred to as **closed exam questions** – a popular form of assessment that is easily evaluated, but has been found to be suitable for assessment of *declarative knowledge* only

multistructural answer – a class of answers containing several parallel *themes*, representing features of the described phenomenon according to the *SOLO* hierarchy. Such answers are more advanced than *unistructural* and less than *relational*. It has been introduced by (Biggs and Tang, 2007)

natural language processing, often abbreviated as *NLP* – the subdomain of computer science that deals with interpretation of written text.

non-player character, often abbreviated as *NPC* – see *virtual character*

object of learning – the phenomenon being studied. Three articulations could be considered according to (Marton et al., 2004), the *intended object of learning*, the *enacted object of learning* and the *lived object of learning*

offender – the *attacker* committing a crime in *crime prevention*

offender presence in situation – one of the causes of crime related to both situational and offender-oriented factors, according to CCO (Ekblom, 2011a)

ontology – a study or representation of the entities in the world. Philosophically addresses entities in the world, but for the purposes of learning and game design addresses the entities that are being explored and studied, the ones not represented considered for not relevant to the explored context

ordinal – data that is defined over an ordered mathematical space

part-task sequencing – sequencing learning materials about a complex subject, starting with learning partial simplified models of the subject in contrast to *whole-task sequencing*. Further explanation and references can be found in (van Merriënboer and Sweller, 2005)

participatory design – a design activity in which end-users contribute to the design of the developed learning tool

performative knowledge – see *procedural knowledge*

persuasion – the process of convincing someone towards a change of behaviour or attitude. It is discussed in (Fogg, 2002)

persuasive games – games as a special case of *persuasive technology*, very similar to *advert-games*

persuasive technology – see *captology*

phenomenographic learning study – see *learning studying*

phenomenography – a qualitative research methodology that investigates different ways in which people experience or think about certain phenomena. In other words phenomenography is the science of studying the relationships between actors and a phenomenon. It has been introduced in (Marton, 1981)

position – the declared goals of one side in a negotiation. This has been contrasted to *interest*, which represents the implicit goals. These two are discussed in (Fisher and Ury, 1991)

possessive knowledge – see *declarative knowledge*

potential threshold concept – A provisionally accepted *threhold concept* until further evidence. Key to the study of threshold concepts because of the sophistication of establishing a threshold concept with high certainty.

presence – along with *absorption*, *presence* and *immersion* psychological construct representing engagement. These constructs are compared in (Brockmyer et al., 2009)

prestructural answer – a class of answers not coherently containing a *theme* that would represent features of the described phenomenon according to the *SOLO* hierarchy. Such answers are the most basic ones, with insufficient complexity to even become *unistructural*. It has been introduced by (Biggs and Tang, 2007)

preventer – see *crime preventer*

proactive knowledge – see *conditional knowledge*

problem analysis triangle – see *crime triangle*

problem-based learning, often abbreviated as *PBL* – a pedagogical principle focusing on problem-solving as a main learning technique. It is related to *learning by doing* and *situated cognition*. Further explanation and references can be found in (Hmelo-Silver et al., 2007)

procedural knowledge, also referred to as **imperative** or **performative knowledge** – a form of knowledge that allows people to perform certain activities. It could be explained as know-how, as opposed to know-what. Further explanation and references can be found in (Gijbels et al., 2005)

propositional knowledge – see *declarative knowledge*

provisional knowledge – knowledge that is understood as the best available explanation of a phenomenon, and due to be replaced with more accurate one once developed.

rapid prototyping – a type of *agile development* in which prototypes are released on a frequent basis.

readiness to offend – possible inclination of the offender to commit a particular crime. One of the offender-oriented causes of crime, according to CCO (Ekblom, 2011a)

recommender systems – computer systems that are developed to provide recommendations based on previous knowledge of the user. Typically employ data mining and machine learning algorithms

reflective practitioner – (Schön, 1991)

relational answer – a class of answers, not only containing, but also developing relationships between several parallel *themes*, representing features of the described phenomenon according to the *SOLO* hierarchy. Such answers are more advanced than *multistructural* and less than *extended abstract*. It has been introduced by (Biggs and Tang, 2007)

resources for crime – various material tools, skills or interpersonal relationships that make it easier for the offender to commit the crime. One of the offender-oriented causes of crime, according to CCO (Ekblom, 2011a)

resources to avoid crime – mostly cognitive capabilities of the offender that could help him/her not to commit the crime. One of the offender-oriented causes of crime, according to CCO (Ekblom, 2011a)

role model – the use of someone's behaviour as a baseline model for comparison to the self-concept. Elaborated in the context of professional development in (Shuval and Adler, 1980)

separation – along with *contrast*, *generalisation* and fusion, one of the four techniques to experience aspects of learning variation. Involves changing a feature of variation while keeping the other features invariant. It has

been introduced by (Marton and Pang, 2006)

serious games – games that have been designed for a purpose different from entertainment. Examples of such games are *learning games*, *art games* or *advert-games*

serious game for learning – see *learning games*

simulation – an interactive environment that models a subset of reality. It has been identified to have four key attributes: (i) a model of reality defined as a system; (ii) a dynamic model; (iii) a simplified model; and (iv) a model that has fidelity, accuracy and validity. Further explanation and references can be found in (Sauvé et al., 2007)

simulation game – a product that combines the typical features of *games* and *simulations*. Further explanation and references can be found in (Sauvé et al., 2007)

situated cognition or **situated learning** – a theory about learning that posits that knowing is inseparable from doing. Researchers in situated cognition argue that context is an important aspect of knowledge and thus abstract learning is an inefficient practice. It has been introduced by (Brown et al., 1989)

situated learning – see *situated cognition*

situational knowledge – see *conditional knowledge*

social awareness stream – a type of social network, usually associated with *Twitter* and *Facebook*. It is typical for such networks that users have networks of friends that they follow virtually. Further explanation and references can be found in (Naaman et al., 2010)

social constructivism – a theory that is a collective perspective on *constructivism*. It posits that knowledge is not being transferred from one person (e.g. teacher) to another (e.g. student), but rather it is being collectively constructed by a community. This holds both for the invention of new theories and for the development of an understanding of already existing cognitive constructs

social motivation – the need people have to interact with other people, be it as a form of *socialising*, *relationship building* or *teamwork*. Along with and *immersion* and *achievement*, this is one of the three main groups of motivational drives to play massive multiplayer online games according to (Yee, 2005)

story world or **backstory** – the narrative environment that provides context to the evolving events. Further discussion is available in (Crawford, 2004)

structure of observed learning outcomes – a hierarchy for the purposes of assessment of written open answers. It features levels from *prestructural*, *unistructural* and *multistructural* to *relational* and *extended abstract*. It has been introduced by (Biggs and Tang, 2007)

summative assessment – the assessment conducted at the end of a learning experience to provide an overall mark of the learner's progress. Further explanation and references can be found in (Atherton, 2013b)

surface learning – learning that focuses on memorising what was directly observed. Contrast this to *deep learning*, where a deeper understanding of the phenomenon is sought, trying to understand how it interacts with its environment. Further explanation and references can be found in (Atherton, 2013b)

target enclosure – the surrounding of the *target person or property*, often with the function to protect it by allowing legitimate access and restricting unintended access. One of the situational causes of crime, according to CCO (Ekblom, 2011a)

target person or property – the possible target (or victim) of an attack, which on its own plays a role (even if passive) in the crime situation. One of the situational causes of crime, according to CCO (Ekblom, 2011a)

table-top role-playing game, also sometimes confusingly referred to simply as **role-playing games** – a traditional form of role-playing game that does not rely on computer support. Games of this kind are very diverse, some of them relying on writing down scores (hence pen-and-paper), others on the presence of a boards (hence board game) and usually involve various game attributes like board (called **board games**) cards (called **card games**), dice (called **dice game**), paper-based scoring (called **pen-and-paper games**) or various combinations of these (hence requiring a table). Further explanation and references can be found in (Edwards, 2001)

technology-enhanced learning – a field within *educational research* dealing with learning with the help of technologies

threshold concept – a form of *troublesome knowledge* that serves as a conceptual gateway in that once grasped, it enables wider understanding. Threshold concepts have 5 typical characteristics. They are *transformative*, probably *irreversible*, *integrative*, *bounded* and are likely to involve forms of *troublesome knowldge*. Further

explanation and references can be found in (Davies, 2003)

toolkit – interactive and responsive environment, consisting of several partially coupled activities

transactional inquiry – an iterative approach of refining the understanding and teaching of a discipline. Further discussion and references can be found in (Zeek et al., 2001)

transformative knowledge – one of the five common characteristics of *threshold concepts*. It reflects the fact that grasping a threshold concept involves both an *ontological* and a conceptual shift. Further explanation and references can be found in (Meyer and Land, 2006a)

troublesome knowledge – a form of knowledge that is difficult to grasp. It is typical for *threshold concepts* that they are closely related to troublesome knowledge. Further discussion and references can be found in (Davies, 2003)

unistructural answer – a class of answers containing one specific *theme*, representing a feature of the described phenomenon according to the *SOLO* hierarchy. Such answers are more advanced than *prestructural* and less than *multistructural*. It has been introduced by (Biggs and Tang, 2007)

users – could mean either *software users*, or *security policy users*, depending on the context of use. In this thesis the former is meant, unless explicitly stated otherwise

virtual character, **virtual actor** or **virtual human** – a computer-generated character that has human-like appearance. Virtual characters are central to research in *interactive storytelling*. Further discussion and references can be found in (Ruttkay and Pelachaud, 2005)

virtual world – an online virtual environment, which is often used as an online communication tool. In contrast to *games*, virtual worlds do not have specific goals. Further explanation and references can be found in (Aldrich, 2009a)

viva – an oral assessment technique that often takes the form of an interview with the examinee. This, along with other assessment techniques is reviewed in (Atherton, 2013b)

whole-task sequencing – sequencing learning materials about a complex subject, starting with learning simplified models of the whole subject in contrast to *part-task sequencing*. Further explanation and references can be found in (van Merriënboer and Sweller, 2005)

wider environment – in crime prevention the environment in which the attack takes place, surrounding all actors in the crime. This is one of the situational causes of crime, according to CCO (Ekblom, 2011a)

within-subjects design – along with *between-subjects design* one of the basic standard techniques to conduct experimental studies. In within-subjects design, participants are split in two or more groups and are subjected to the same interventions in the same setting, but in different order. For a close observation, assessment measures are administered before and after each intervention. This is done as a way to ensure that any observed differences between the groups are an effect of different sequencing between groups.

worked example, also sometimes referred to as **worked-out example** or **working example** – a learning techniques that involves demonstrating a solution to learners. Further explanation and references can be found in (Sweller, 2006)

working memory – a capability of the human brain to process information for learning and other complex cognitive tasks. Because of its limited capacity research in *cognitive load theory* is looking for ways to optimise its use. Further explanation and references can be found in (Baddeley, 1992)

Appendix A: Interactive Environments

Reviewed interactive technologies for learning for the purposes of the state-of-the-art overview, provided in Chapter 3. Two reviews are included – one on project management computer games, and one on a broader, but smaller pool in information security, including paper card games.

Project Management

Countdown –
http://www.paradigmlearning.com/subpages/Countdown_A_Strategy_Game_for_Project_Teams.asp

Sharkworld – http://www.sharkworld.nl/

SDSU Project Management Simulator/e-Game –
http://edweb.sdsu.edu/people/etapia/program_mgmt/index.htm

TOPSIM – Project Management – http://www.tatainteractive.com/TOPSIM/TOPSIM_Project_Management.pdf

SimulTrain – http://www.quahance.com/simulations.html
http://www.hemsleyfraser.co.uk/LearningSolutions/TechnologyEnhancedLearning/ProjectManagementSimulation
http://www.tenstep.com/open/miscpages/90.6.PM50.00PMSimulation.html
http://www.sts.ch/index.php?option=com_content&task=view&id=138&Itemid=174
http://en.wikipedia.org/wiki/SimulTrain

CalTrans Project Management Simulation – http://www2.dot.ca.gov/hq/cpsd/PM_sim/

Requirements Game – http://www.clei.cl/cleiej/papers/v10i1p3.pdf

Family Life – http://www.newhabits.co.uk/family-life-try-our-project-management-challenge

Synergest's Project Management Simulation – http://www.synergest.com/project-management-simulation.htm

Race to Results Project Management Simulation – http://h10076.www1.hp.com/education/race-to-resultstest1.htm

Leadership-in-Action Project Management Simulation – http://forio.com/lead.htm

PROTEST - http://www.simulations.co.uk/FUNCTN.HTM#PROTEST
http://www.simulations.co.uk/download/PROTEST.pdf

Harvard's Project Management Simulation – http://hbr.org/product/project-management-simulation-scope-resources-sche/an/3356-HTM-ENG
http://forio.com/simulation/harvard-project-management-demo/login.htm

Celemi Cayenne – http://www.celemi.com/What-we-do/Business-Simulations/Celemi-Cayenne/index.php
http://www.ekgroup.co.uk/content/tools/cayenne.asp

G2G3 Polestar Project Management Simulation – http://www.beyond20.com/polestar-pm.php
http://www.youtube.com/watch?v=iP8z85YT5QA

StevensTech Project Management Capstone Simulation – http://webcampus.stevens-tech.edu/project-management-simulation.aspx

Adam Montgomery's Project Management Simulation – http://www.quanta.co.uk/news/2010/05/first-run-our-project-management-simulation-tomorrow

PROSIGA Project Management Simulation Laboratory – http://www.informaworld.com/smpp/36433788-61759723/content~db=all~content=a713842593

ManSims Project Management Simulation – http://www.mansims.com/PMworkshop.html

Paper Planes – http://www.strategicdevelopment.com/simulations.php?simulations_id=5

CBT Module Simulation and Scenario – http://www.citeulike.org/user/mapto/article/7512686

Contract & Construct – http://www.citeulike.org/user/mapto/article/7499944

Prendo's Management Games – http://www.prendo.com/simulations/projectleadership
http://www.prendo.com/simulations/leadingchange
http://www.prendo.com/simulations/managingprojectteams
http://www.prendo.com/simulations/projectcontrol
http://www.prendo.com/simulations/managingstakeholders

SimProject – http://www.fissure.com/Project_management_workshops.cfm
http://www.youtube.com/watch?v=O8WlO94wtkE

Project Team Builder – http://www.springer.com/engineering/production+eng/book/978-1-4419-6462-5

Vivian Valiant and the Crown of Coaching – http://www.enspire.com/simulations/mastering_management

4D virtual construction technology – http://atmae.org/jit/Articles/park010407.pdf

AXL – Army Excellence in Leadership – http://ict.usc.edu/projects/axl_army_excellence_in_leadership/

VOLT - Virtual Officer Leadership Trainer – http://ict.usc.edu/projects/volt/

Second Life – http://www.citeulike.org/user/mapto/article/7509825
http://www.gepros.com/vtr/
http://second-life-elearning.ahg.com/training_simulations/teamwork_tester_1.htm
http://second-life-elearning.ahg.com/training_simulations/second_life_working_under_pressure_simulation.htm
http://solutionproviders.secondlife.com/provider/show/id/1064
http://hmsystems.net/products.html
http://wiki.sla.org/pages/viewpage.action?pageId=11370942
http://www.lancs.ac.uk/postgrad/ryanm2/SLEDcc08_ryan_paper.pdf
http://www.youtube.com/watch?v=CWfvqkkk0yM
http://www.youtube.com/watch?v=E3D6gN2j_48
http://www.angellearning.com/products/secondlife/downloads/The%20Power%20of%20Virtual%20Worlds%20in%20Education_0708.pdf

Teleplace – http://www.teleplace.com/
http://www.teleplace.com/solutions/workgroup_training.php
http://www.teleplace.com/solutions/workgroup_training.php (p18)

MPK20 – http://labs.oracle.com/projects/mc/mpk20.html

The Medici Game – http://www.celemi.com/What-we-do/Business-Simulations/The-Medici-Game/index.php

Project Risk Board Game –
http://www.successfulprojects.com/Store/ProjectRiskBoardGame/tabid/68/Default.aspx

Vleader – *see Appendix C*
http://www.simulearn.net
http://www.youtube.com/watch?v=Qy3TLyj-HCY
https://www.youtube.com/watch?v=YEZHhYHW7JM

Civil Engineering Project Management Game – http://www.citeulike.org/user/mapto/article/7492599

Baird's Project Management Game – http://www.citeulike.org/user/mapto/article/7492664

SOFTSIM – http://www.systemdynamics.org/conferences/1992/proceed/pdfs/barla059.pdf

Project Management Simulation Training Game – http://cedb.asce.org/cgi/WWWdisplay.cgi?8602373

Information Security

CyberCIEGE – http://www.cisr.us/cyberciege/

Elevation of Privilege – http://www.microsoft.com/security/sdl/adopt/eop.aspx

Ctrl-Alt-Hack – http://www.controlalthack.com/

Privacy Game – http://www.vome.org.uk/privacy-game/

Appendix B: vLeader Study

Data about the level of engagement of students is provided in the table on page 233. And student responses to the test based on a Likert-scale are provided in the table on page 234. Note that samples of the open questions were already provided in Chapter 9. The sets provided here are not the complete datasets, rather randomly sampled examples to illustrate the data.

seudonym	condition	scenario1	scenario2	scenario3	scenario4	scenario5
P03	control	2	3	0	0	0
P06	control	2	2	5	3	0
P08	control	4	0	0	0	0
P09	control	7	4	10	5	0
P12	control	14	5	8	4	4
P20	control	2	2	2	4	5
P33	control	19	9	5	8	4
P38	control	10	7	9	3	4
P40	control	1	2	2	2	1
P46	control	2	4	3	0	0
P49	control	10	6	6	0	0
P52	control	3	2	4	0	0
P56	control	4	1	3	1	1
P05	experimental	13	2	4	2	0
P18	experimental	2	2	5	1	0
P22	experimental	16	2	7	0	0
P24	experimental	5	2	2	1	0
P25	experimental	3	2	0	0	0
P27	experimental	2	2	2	1	1
P34	experimental	4	2	2	4	0
P41	experimental	34	13	14	10	4
P44	experimental	3	0	0	0	0
P53	experimental	15	9	5	3	4
P55	experimental	30	3	4	3	2
P01	no-play	2	1	1	0	0
P11	no-play	0	0	0	0	0
P15	no-play	1	1	1	0	0
P19	no-play	0	0	0	0	0
P23	no-play	0	0	0	0	0
P26	no-play	0	0	0	0	0
P32	no-play	0	0	0	0	0
P43	no-play	1	1	0	0	0
P47	no-play	2	1	1	1	0
P57	no-play	0	0	0	0	0
P58	no-play	2	1	0	0	0
P59	no-play	0	0	0	0	0
P60	no-play	0	0	0	0	0

pseudo-nym	condition	Q4 – before							Q4 – after						
		I	ii	iii	iv	v	vi	vii	I	ii	iii	iv	v	vi	vii
P01	control	1	5	2	1	3	1	1	1	2	5	3	1	1	1
P06	control	4	5	3	1	3	1	1	2	5	5	1	2	1	3
P08	control	3	4	4	2	1	1	2	3	4	4	3	2	2	3
P10	control	2	3	3	2	2	2	2	2	2	3	2	1	1	2
P12	control	5	5	5	1	1	2	1	1	5	5	1	1	2	2
P14	control	4	5	5	2	3	3	1	1	3	5	2	1	5	1
P31	control	3	4	3	2	2	1	1	2	1	4	2		2	1
P33	control	4	4	4	3	3	2	4	3	2	5	4	2	4	2
P35	control	4	3	3	2	2	2	3	2	3	2	3	2	3	3
P38	control	2	4	4	2	1	1	1	2	1	3	1	1	1	1
P47	control	1	5	3	2	1	1	1	2	1	2	1	1	1	2
P49	control	1	4	2	1	1	1	1	1	5	5	1	1	1	1
P50	control	4	3	5	1	1	2	1	3	2	5	1	1	1	2
P52	control	3	2	5	2	2	2	2	3	2	4	1	1	1	4
P54	control	2	1	1	3	4	2	2	2	2	3	3	3	2	3
P02	experimental	5	4	4	4	2	4	3	5	3	4	2	2	3	5
P13	experimental	5	2	4	2	2	4	2	2	3	5	4	2	2	2
P16	experimental	3	5	5	2	2	1	1	3	2	5	3	1	2	3
P17	experimental	5	5	5	3	1	1	1	5	1	5	3	1	1	1
P18	experimental	5	5	5	3	1	1	1	5	3	5	1	1	1	1
P24	experimental	1	3	2	4	1	3	3	1	4	3	2	1	3	3
P25	experimental	5	5	1	3	1	1	1	5	3	5	1	1	1	3
P28	experimental	1	5	2	1	2	1	1	1	1	5	1	1	1	1
P34	experimental	2	3	4	2	3	1	1	1	3	2	2	4	1	2
P41	experimental	2	4	5	3	1	1	2	3	1	5	1	2	3	2
P44	experimental	2	1	4	3	2	3	2	2	2	4	2	1	3	2
P48	experimental	5	5	3	1	5	1	1	5	3	1	1	1	1	1
P53	experimental	3	1	3	2	2	2	2	3	1	3	2	1	1	1
P55	experimental	5	1	5	1	3	2	2	5	5	5	1	1	1	1
P07	no-play	2	4	4	1	1	1	3	4	1	4	2	1	2	2
P11	no-play	4	2	5	1	2	1	1	1	1	5	3	2	2	5
P19	no-play	5	4	2	2	1	3	1	5	4	3	5	1	4	4
P51	no-play	2	4	5	4	2	5	5	2	1	5	1	1	2	2
P57	no-play	5	1	5	3	1	1	1	5	3	5	4	2	2	2
P59	no-play	5	2	5	3	1	1	2	5	3	5	2	1	1	2
P60	no-play	3	5	5	3	2	3	2	4	4	5	3	2	3	4

Appendix C: CCO

A sample of the project development list of tasks is provided in a table on page 236.

The full game experience is available at http://orpheus.cs.ucl.ac.uk/cco-is

Scenario: Meltdown

A big IT consulting company prides itself on its flexible software development service and has been contracted to undertake a number of big projects for leading global brands. The bulk of development work in the company is being done by individual employees in its main open-plan back-office.

The company has moved its office internationally between tax-havens in Europe. In 2009 two thousand people were offered relocation to another country. More than half of them signed up, but after some 18 months there were barely three hundred remaining. The relocated staff were not the only ones to leave, the overall average period of employment also fell to 10 months, from 31 in 2007. There was a general feeling in the office of being underpaid compared to what is being delivered.

Although HR has managed to supply a steady flow of skilled recruits, the turnover meant that more people had access to work with customer's sensitive data. The turnover also resulted in reduced quality of delivered software services, with a number of cases (see text to the right) of intentional installation of malware. These involved both software developers and system administrators, the latter responsible for monitoring security systems. Because of frequently changing staff, a strong corporate culture cannot form. The problem reached the extent that office jokes spread about making the big hit before leaving to work with another company in the industry. Such stories might have attracted new recruits driven by criminal intents.The company maintains a policy of keeping silent about attacks and insider stories talk only about successful attacks. A case of end-user data being leaked in public turned into a media scandal with journalists attempting to investigate previous cases.

Management wants to reduce sabotage and IP thefts and improve the image of the company: both internally and externally. As a wider trend long-term customers have raised concerns, new customers have become wary and slow to place new orders.

A culture of viewing the company as an evil empire has developed among employees. Management estimates that the level of planted backdoors could amount to up to 5 times more than the disclosed cases.

Cases of Insider Attacks

1. A number of cases of activation of anti-virus security mechanisms because insiders attempted to deploy off-the-shelf malware.
2. There was also a case of blackmail received about software vulnerabilities that would be triggered if a sum of money is not paid.
3. In one particular case, a group of two developers and two administrators leaked onto a publicly accessible website personal data of 500,000 end users of a key customer. Suspicions remain that this attack was ordered by a competitor.
4. Another attack resulted in 40% of the company's in-house data servers being down for 50 minutes which resulted in customer claims for £58,000.
5. One incident leaked company financial data, revealing what percentage of income goes to back-office salaries, which further undermined employee morale.
6. In one of the cases a group of five managed to leak contract data and use it to win over a customer for their newly established company.

CCO Lab Study

On page 238 are provided user-generated causes extracted from the toolkit database. Similarly page 239 is for such interventions, and page 240 – for peer assessments. Page 241 contains a sample of the expert assessment of ideas, suggested in the pre- and post-tests of Study 2

CCO Class Studies

Second Scenario: Spoonlure

A leading multinational retail bank has seen the number of phishing attacks on its employees and customers increase exponentially, almost doubling annually. The more disturbing evidence within this emerging trend is attackers focusing on credentials of key bank employees that would allow them to distribute phishing content that appeared to come from legitimate sources.

Attack attempts have been detected in most of the 28 countries that the bank operates in, but the US arm of the company registers an average of 43 spear phishing or whaling e-mails per day.

Attacks would typically start with prolonged campaigns of collecting personal information about bank employees of interest. This preparatory phase typically culminates in a personalised scam which aims to acquire the target employee's credentials to access a system in the intranet that would provide the necessary privileges to organise a one-time mass phishing campaign. In particular, the campaign would focus on circumstances where the bank finds target customers are difficult to identify and thus to inform of the breach personally. The attackers rely on this and the reluctance of the bank to contact all customers to inform them about a possible breach that would have only affected a relatively small portion of them. Different steps of such attacks could be undertaken by different parties, with sophisticated anonymous online markets where intermediate results (e.g. the target employee's spersonal details) are exchanged.

Typically, target employees are staff who have occupied an important position in the bank for a couple of years, when their initial alertness to crime threats starts to wear off (e.g. due to lack of incidents experienced), and habits related to security hygiene have not developed strongly.

A cluster of attacks have been identified as originating from various groups in a small town in Eastern Europe dubbed Hackerville, where tricks and techniques related to hacking are widely accessible public knowledge. Any lack of own expertise (e.g. when necessary coding, design, logistics and translation) is usually overcome through exploiting contacts within legitimate professional communities, and subsequent outsourcing to members of those communities who are more willing to take risks in search for better returns.

Spearphishing takes place via all possible media – most commonly e-mail, phone or websites, but even face to face. It is estimated that the average loss resulting from such two-staged spear phishing attacks is $150,000. However, the far bigger negative impact is on the corporate image, with 42% of customers declaring that they are less likely to do business with companies that they know are targeted by phishing attacks.

This scenario features two subsequent and distinct phishing attacks – the first spear phishing targeting the employee, and the subsequent mass phishing on customers, originating from the compromised account. Please focus on possible ways to address the first.