Teaching a Complex Cognitive Skill Using an Educational Game

How non-game related design methods help structure educational game design

Ivo Bril
2001179
November 2016

Master's Thesis
Human-Machine Communication
University of Groningen, the Netherlands

Internal supervisor:
Prof. Dr. Niels Taatgen (Artificial Intelligence, University of Groningen)

External supervisor:
Dr. Nick Degens (Hanze University of Applied Sciences)

university of groningen / faculty of mathematics and natural sciences / artificial intelligence

Abstract

How does one teach a complex cognitive skill? Although there are several educational games whose aim is to teach these skills, there is little literature on how to teach it using an educational game. There are, however, several non-game related methods for teaching complex cognitive skills. These methods have been addressed as possible tools for educational game designers to use, yet have only received little attention.

In this thesis, we will use two of these methods in combination to help formalize the design process of an educational game. A cognitive task analysis is combined with the four component instructional design method to explore a domain, identify a cognitive skill in need of training, construct a general training plan, and implement this by changing an existing game. Finally, we test to see whether playing the game causes participants to become better at the identified cognitive skill.

To find out whether a complex cognitive skill is trained more effectively using feedback that is more steered towards reflection, two versions of the game were made: one with summative, non-reflective feedback and the other with formative reflection-inducing feedback.

Results show that participants did not get better at the game and that there was no difference between the formative feedback group and the control group. The non-game related methods show promise for formalizing the design of educational games and are discussed thoroughly.

Acknowledgements

I would like to thank Nick Degens and Niels Taatgen for their effort and patience in guiding me through this project. Their feedback, insights, and knowledge really helped me grow during my final year at the University of Groningen.

I would like to thank the Hanze University of Applied Science and the University of Groningen for giving me the opportunity to attend conferences which accepted papers that I wrote from research done for this thesis. I also want to thank Eelco Braad for joining in on the discussions between Nick and me, shedding his light on the subject or just to have a good laugh.

I want to thank my family, who was always eager to hear about the developments of my thesis and the conference papers.

Lastly, I want to give a special thanks to Eveline Broers, who has been my most critical and brutally honest reviewer, as well as the one who has kept me going when times were hard.

Contents

1. Introduction

How can technological advances be used to help humanity? This question is at the centre of many fields that aim to aid humans through technical means. For many of these fields the result is already visible: smart phones are used throughout our daily lives, cars are used to travel large distances each day, not to mention the small things such as sunglasses being a result of space travel research.

For fields such as education, technological advances have always been met with enthusiasm as well as careful scepticism. Even the earliest of media, such as simply presenting information through imagery or the use of radio, was examined to see how it could be used to educate (for a review see: Reiser, 1987). However, results on effectiveness were mostly inconclusive (Wilkinson, 1980). Only a little more than a decade ago even the use of computers in education received harsh criticism (Cuban, 2001). Since then, the computer has gained a stronger foothold in schools, allowing a new kind of technological advancement to arise: educational games.

What are educational games?

Educational games are games with more diverse purposes than only providing entertainment. An educational game aims to replicate the engaging factors of entertainment games, in an effort to teach the user, change the user's behaviour, or to reach other non-entertainment goals. The big difference between entertainment games and educational games is their use for the player in the real world. An entertainment game focuses on providing a session of entertainment; the user is not expected to apply anything learned in-game to the real world. For an educational game, playing the game serves goals that are also relevant outside of the game's environment. These games for example aim to change unhealthy behaviour[1], teach a range of subjects from mathematics[2] and physics[3] to topics such as scientific reasoning[4], or even to recruit potential soldiers[5]! These games use a playful environment to engage and motivate its players as a means to reach a goal other than entertainment alone.

The playful environment allows the user to learn and train skills in a safe environment in which failures are learning moments and problems are perceived as engaging challenges. Small, achievable goals combined with just-in-time information provide the user with clear directions and a feeling of empowerment, as the goals are at the edge of the user's range of capabilities (Gee, 2003).

Why would educational games work?

If previous digital media such as television and radio turned out to be less effective at educating students than traditional methods (Wilkinson, 1980), then why would educational games be able to? It is true that empirical research does not provide a conclusive answer on the question whether educational games are more effective at educating their users than more traditional methods (Young et al., 2012).

However, a game has several innate traits which lend themselves well as an educational tool (as opposed to the other media such as television). One of them is their *interactive nature*. If literature on learning has taught us one thing these past two decennia, it is the fact that a student learns more when he or she is actively interacting with the to-be learned knowledge (e.g. Bransford, Brown, & Cocking, 2000).

---

[1] http://www.hopelab.org/innovative-solutions/re-mission/ (Kato et al, 2008)
[2] http://dragonbox.com/home
[3] https://sites.google.com/site/surgeuniverse2/home/game/fuzzy-chronicles-1 (Adams & Clark, 2014)
[4] http://questatlantisblog.org/ (Barab et al., 2009)
[5] https://www.americasarmy.com/

Students can learn more complex skills and subject matter under the guidance of a more knowledgeable person than they would have been able to on their own. The level of knowledge learned this way is often referred to as being in 'the zone of proximal development', and providing the student with the guidance to learn this knowledge is known as 'providing scaffolding'. This concept was introduced in the early 20th century by Vygotksy (1978). The idea of personalized guided learning (having the curriculum adjusted to the student) has expanded and proven itself in other aspects of learning as well (e.g. feedback: Shute, 2008). To be able to provide such personalization, the teacher needs to know where the student is in terms of prior domain knowledge. As games require the user to interact, the information provided by this interaction shows the user's prior knowledge and can be used to personalize the game's teaching. While this feature could be implemented in media such as television and radio, it is at the heart of good game design.

The above two aspects (the interactive nature and the proximal development applicability) are not unique to games however, as tutoring systems provide such an experience as well. Interestingly, the two fields have a lot in common in terms of goals; both wish to educate their users. Although the distinction could be made that a tutoring system provides education in a more traditional set-up, such as providing personalized math questions to teach math, even that distinction is blurred by the increasing interest in narrative as instructional method (e.g. McQuiggan et al., 2008). This increasing overlap makes it harder to fully argue where the field of intelligent tutoring systems ends and the field of educational game design begins.

If there is one way to distinguish between both types of software, it would be the user's underlying intent. An educational game will be more focussed on appealing to the play- and engagement-seeking behaviour of the user, whereas a tutoring system appeals more to the need to learn (even though both kinds have the goal to educate their users). Thus, an educational game would be more appealing to students during their spare time as opposed to a tutoring system, while a tutoring system in turn would be more appealing during school.

Still, the notions of interactivity and applicability of proximal development give an idea of how a game environment could be used to educate and why it could be good at doing so. The following section will briefly describe the state of the field of research on educational games; its accomplishments and the problems that lie ahead.

What is the state of the field?

The art of designing educational games that provide an optimal learning environment has been an important topic of research for over two decades now. Although the effectiveness of games as educational tools was heavily disputed over the past decade (Vogel et al., 2006; Ke, 2009; Clark, Yates, Early, & Moulton, 2010), big strides have been made in those years as well. One of the most notable is the fact that the field of research has produced good examples of effective educational games (Barab et al., 2009; Habgood & Ainsworth, 2011).

The fact that games *can* be an effective learning tool raises a follow-up question: what makes an effective educational game actually effective? This question is at the heart of contemporary research in educational game design (Young et al., 2012; Clark, Tanner-Smith, & Killingsworth, 2015),

Educational game design research is now geared towards finding the reasons *why* certain educational games are effective while others are not. The goal of this direction is to provide insight in what a game needs to be *educational*, in an effort to give designers of educational games the knowledge on how to make an effective educational game.

To make results of research usable for designers, researchers have tried to bridge the gap between researchers and designers by developing overarching taxonomies and frameworks (e.g. Arnab et al., 2014; Carvalho et al., 2015). This work has paved the way in which educational games can

be described and analysed objectively. It also allows researchers to get closer to underlying principles of effective educational games by showing overlapping design choices of different educational games.

However, one aspect that is still lacking in this formalization process is a cognitive perspective on educational game design, as most of the formalization has come from pedagogical theories (e.g. Gunter, Kenny, & Vick, 2008). By a cognitive perspective on educational game design, we mean using what is known about how people process information, recall information, learn new skills, refine learned skills, etc. to determine how an educational game should be designed. This perspective looks at moment-to-moment game play as well as the overall structuring of the learning experience in order to validate game design choices on their effect on achieving the game's learning goals.

While some research has tried to link game elements to cognitive *learning outcomes* (Wilson et al., 2009), very little research has looked into the role of cognition during the *process* of playing an educational game (e.g. Greitzer, Kuchar, & Huston, 2007). Knowing the consequences of design choices on the learning process of the user helps us in answering the *why* question posed above (Clark, Tanner-Smith, & Killingsworth, 2015). We believe that focussing on the user's cognition can help pre-determine (at least to some extent) whether a design choice will harm (or help) the user's learning process.

An example of the impact of cognition on the effectiveness of an educational game is found in a paper which focuses on the interactive elements of a game and their effect on the user (Tawfik, Moore, He, & Vo, 2012). The 3D game that the participants played was met with confusion by participants with less gaming experience. The confusion was not caused by the difficulty of the domain, but by the gameplay itself. Controlling and navigating the avatar through a 3D environment required too much attention of the less experienced participants causing them to be unable to learn the domain knowledge presented by the game. In this example, the less experienced participants had to learn these control and navigation skills first, before any domain knowledge could be processed; it showed the influence cognitive processes have on the effectiveness of an educational game.

How do we systematically involve cognition in our design?

Even though there are some papers, such as Greitzer et al.'s (2007) work, that provide some general guidelines based on cognitive principles, there is still a gap between knowing how cognition works and structurally embedding it in the design of an educational game (Bellotti, Berta, and De Gloria, 2010).

Taking a step back from this difficult problem, a question rises up in regards to the other media we discussed. How have other educational tools tackled the problem of working with the cognition of the learner? And what techniques did they use? Perhaps we can learn from or even use non-game related methods that address these issues and apply them to ours? This would not only help to properly embed a cognitive perspective in the educational game design process, but also formalize the design process as a whole.

This thesis will look into the applicability of non-game related educational and instructional design methods to see whether they can structurally embed a cognitive perspective into the design of an educational game.

To do so, some questions have to be answered:

- What is the relationship between the elements of an educational game and the cognition of the user and how do they relate?
- What non-game related methods should we use to train the identified skill?

- How do we use these methods to structure an educational game's design process to include a cognitive perspective?
- Does including this cognitive perspective in an educational game's design process lead to a more effective learning experience?

We expect that answering these questions (even partially) will tell us something about the underlying principles of effective game design, the role of cognition, and how to incorporate cognition into a game's design.

Thesis structure

In the next chapter, we will provide an answer for the first question using literature from the fields of cognition, learning, instructional design, and educational game design. Chapter 3 will introduce a problem domain in which we can instantiate the research done in the previous chapter. Chapter 4 describes the additional research done to properly understand the domain and skill the game will teach. Chapter 5 will describe an existing prototype and the non-game related method used to guide the design of the adaptations, giving us an answer on the second and third question. The setup of the experiment is then described in Chapter 6. The results gained from running the experiment are presented in Chapter 7 (answering the fourth question) and will be discussed in Chapter 8 alongside our experiences with the non-game related methods and general insights. Lastly, Chapter 9 will conclude this thesis with some ideas for future research.

2. Background Literature

"What is the relationship between the elements of an educational game and the cognition of the user and how do they relate?"

The above question will be answered in this section by identifying and linking two things:

1. the elements that are found to be critical for learning in educational games;
2. the role of cognition in these elements as extracted from literature on cognition and instructional design.

Game elements

Garris et al. (2002) were the first to objectively identify a set of eight game elements having a significant effect on learning. Wilson et al. (2009) used these game elements for their research on cognitive learning outcomes and expanded on it by identifying ten new elements. However, these eighteen elements were later revised as there was too much overlap between them (Bedwell et al., 2012). This resulted in an overarching taxonomy of nine categories, which categorize the eighteen elements as identified by Wilson et al. (2009). Table 2.1 provides an overview of the nine categories with a short description.

Table 2.1: Overview of each game element category accompanied by a short description.

Educational game element category	Short description
Action Language	The 'language' in which the player communicates with the system. This includes the interface of the system and the method by which the player interacts with the system (e.g. using a keyboard and mouse to play a game).
Assessment	"Assessment describes the nature and content of any feedback given to the player during the course of a game (e.g., debriefing, feedback, scoring, etc.)." (Bedwell et al., 2012)
Conflict/Challenge	Describes how problems are presented in the game, their (possibly adaptable) difficulty, and how well the problem is defined (e.g. the problem is left vague on purpose by the designers).
Control	"Control refers to the degree of actual interaction and agency the player is capable of in a game." (Bedwell et al., 2012)
Environment	The (digital) representation of the world (and its surroundings) in which the player is immersed.
Game fiction	Describes the story of the game in attributes such as fantasy (elements that do not exist in the real world) and mystery (the gap between known and unknown information).

Educational game element category	Short description
Human Interaction	All human-to-human contact present in a game. This can be defined by the tools with which users can interact with one another inside the game environment and the relationships between users in the game (e.g. cooperative or competitive).
Immersion	Immersion describes how the player perceives and feels about the game's environment and story (i.e. the *Game fiction* category). It is seen in for example the acceptance of an alternate reality and the loss of fear of failure, all of which are influenced by the way the game stimulates the senses of the user (e.g. use of relevant dramatic/novel audio and visual stimuli. Garris et al., 2002).
Rules/Goals	The goals describe *why* the user is interacting with the game. The rules determine how the *Action Language* translates to in-game actions and when the user achieves the goal.

To see how cognition plays a role in each of these elements, we will discuss them from a cognitive perspective one by one. Doing so, we will be able to see where we can apply the literature on cognition. Having this cleared up will help us design our game later on.

Before we discuss each element, there is already one cognitive construct we wish to discuss, as it is present in most of the above mentioned categories: *transfer*.

Transfer

When a learnt skill (or parts of said skill) is applied in a different context than the one it was trained in, the skill is said to be 'transferred' to the novel context (Barnett and Ceci, 2002; Taatgen, 2013).

Education needs transfer; the skills and knowledge learned in school is not meant to only assess whether a student can learn, it is meant to be used by said student *outside* of the school environment in which the knowledge is learnt. The same goes for educational games, as they too need their knowledge to be transferred to situations outside of the game itself, otherwise there is no distinction between entertainment and educational games.

Barnett and Ceci (2002) describe several dimensions (contexts) in which transfer can take place: *Knowledge*, *Physical*, *Temporal*, *Functional*, *Social*, and *Modality*. They also distinguish several aspects of the learned content: *Learned Skill*, *Performance Change*, *Memory Demands*.

These dimensions and aspects allow us to identify a *distance* between the context in which the skill is acquired and the target context in any of the given dimensions. For example for the *Temporal* dimension, transfer is more likely to occur if the learned skill is tested in the target context the day after the skill is learnt (small distance) instead of one month later (large distance).

Transferring on a *small* distance is referred to as *near transfer*, whereas transferring on a *large* distance is known as *far transfer*.

These different aspects and dimensions will be described more thoroughly at each of the game categories, as most categories deal with one or more of the transfer dimensions and learning content aspects.

Action Language

Transfer: Modality

In educational games, the user often interacts with the system using the mouse and/or the keyboard. Depending on the learning goals of the game however, the user will not use those tools to perform the learned skill in the real world context. In such a case, there is a distance in the *modality* dimension.

For more physical training, games have the option of using special hardware to incorporate the actual movements in the game play (Jalink, 2014). This helps to close the distance between the two contexts. For cognitive skills such as decision-making, you can argue that modality is not as relevant. The learned skill is mostly a mental skill, thus not requiring any specific physical way to perform it.

Interface

The interface of the game is also part of the Action Language, and determines where the player's attention is steered towards and how it is managed. As interfaces have been used more often for educational purposes, there is plenty of literature discussing how to manage the user's attention properly. Clark and Mayer (2011) have identified several basic rules on instructional design using multimedia, all of which deal with how to properly manage the user's attention. The

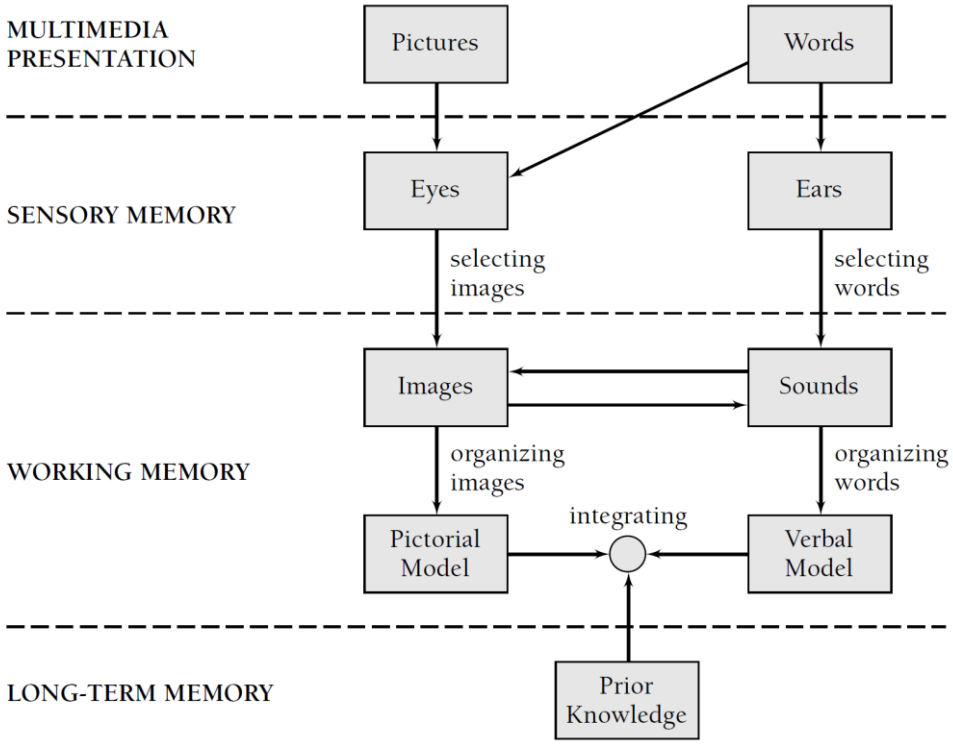


Figure 2.1: A graphical representation of the processes involved in interpreting audio and visual stimuli. Reprinted from Figure 3-2 in Mayer (2001). Copyright © 2001 by Cambridge University

majority of these principles stem from the fact that the human cognition has multiple ways of perceiving (the senses) and that each 'channel' can only process one thing at a time, as seen in figure 2.1 (Baddeley, 1997). This corresponds with the different regions of the brain being responsible for different senses (at least for processing audio and visual stimuli).

An example of one of Clark and Mayer's (2011) principles would be the *modality* principle, which states that learning is fostered when the learning material uses both 'channels' to instruct. For example, it is better to use an animation of how something works whilst the explanation is given through audio narration, than to have an animation be explained by on-screen text. The latter demands the user to divide his or her attention between two visual stimuli, which hampers the learning process (Chandler and Sweller, 1992). This leaves the parts of the brain that process audio stimuli unused, whilst the visual processing parts of the brain have to switch between two stimuli.

In game design, visual and audio stimuli are used almost exclusively (with the occasional tactile feedback). For educational game design specifically, both game play as well as educational content has to be presented through these channels. Thus one should take into account when to present educational content and through which channel, as playing the game itself already means using those channels.

An example of incorrect interface design in an educational game can be found in SURGE, an educational game that teaches Newtonian laws of physics (Clark et al., 2011). The player has to control a space ship in real-time using boosts of force to move, change trajectory, and stop (see Figure 2.2). If the player is unable to control their ship and hits several boundaries, the ship explodes and the player loses.

The problem arises in the way feedback is presented. In case of a collision, a large pop-up is placed in the centre of the screen with feedback on how the rules work. While the feedback itself is good, the game continues in the background, requiring the player to divide attention between the feedback and the ship. The feedback remains visible for only a few seconds as well, making it hard for the player to take the time and interpret it. By simply pausing the game when an error is made and requiring the user to continue the game by clicking a 'close' button, the attention of the player is channelled much more smoothly (barring the consequences this could have for the engagement-factor of the game play).

Figure 2.2: Left image shows a level of SURGE. The goal is to reach the cube in the middle. The right image is a close-up of the ship; the red arrows represent constant force, the yellow arrow represents the speed (length of arrow) and direction of the ship.

Summary

From a cognitive perspective, the element *Action Language* requires us to properly steer the attention of the user to the information that is most relevant at a given time. The dual-channel theory provides insight into how the user's attention can be used optimally to make sure that the user is able to interpret the educational content.

Assessment

Feedback

One aspect of assessing a player is providing clear feedback when an error is made to make sure that the player can learn from their mistakes (Clark et al., 2010; Shute, Rieber, and van Eck, 2011; Hattie and Yates, 2014). In video games, feedback has several forms; the simplest one being the visual and/or audio feedback you receive by interacting with the system itself (e.g. pressing a button to move your avatar). More sophisticated forms of feedback are seen in the loss of valuable resources after making an error, punishing the player to make them behave differently the next time such a problem occurs (as seen in the collision punishment in the SURGE example on the previous page).

When it comes to the educational content, feedback is a strong tool for a game to help the player learn. It can promote a deeper understanding of the educational content if done right, as well as motivate the player and increase their self-efficacy (Clark & Mayer, 2011; Hattie & Yates, 2014). Shute's (2008) article on the role of feedback in learning expresses this in great detail.

One of the most important things is the malleability of feedback to the right situation, right timing and the right recipient. There is a balance to be struck here, as the more complex and lengthy feedback becomes, the more cognitively demanding it becomes to process. Thus, a fast paced game would probably not benefit from extensive immediate feedback. Instead, this can be done after a level (or sequence of levels), when the player has the time to interpret the feedback.

It is important for a designer to keep in mind what the feedback is supposed to achieve. For example, it has been found that providing reflective questions and guidance can stimulate more critical thinking, and in turn help with teaching strategic knowledge (e.g. Wetzstein & Hacker, 2004; Zohar & David, 2008). However, if reflection is not what the feedback is meant to achieve, as it for example needs to train a specific procedure, then one needs to look more into just-in-time feedback aimed at assisting the player in executing the procedure.

Scoring

Games are known mostly for a challenge you can win, and the fun that is had whilst trying to best the challenge. To let the player know how well he or she is performing, a game oftentimes uses some kind of scoring system. In an educational game, it is important for a designer to think of what scores they find relevant. If a player's score is determined by being fast, then speed needs to represent the mastery of the educational content. Otherwise the game rewards players for skills irrelevant to the learning goals, causing players to focus on those skills instead.

An example of this is seen in the SURGE example earlier (Clark et al., 2011). Each level in SURGE has a timer measuring how long the player is taking to clear the level. One of the rewards in SURGE is a score at the end of each level, which is partially determined by the time it takes for a player to reach the finish line; the faster the player finishes a level, the higher his or her score is (barring the other score-determining variables). Doing so encourages the player to finish levels as quickly as possible. This in turn can cause players to fly their ship more recklessly in an effort to get to the finish line sooner, hampering their focus on how the ship is controlled and why.

Transfer: Performance Change

In line with the *scoring* section, the learned content is more easily transferrable if the requirements of the target context (i.e. what the learner is judged upon) are similar to the requirements of the context in which the content was learnt. For example: a hypothetical game only requires the learner to know *how* to calculate a derivative and scores on how fast the learner can do that procedure. If the target, real-life, context requires the learner to also know *when* to calculate the derivative, and scores the learner on how well the appropriate moments are recognized, there is a gap between expectations. This difference will possibly leave the learner frustrated, as the game failed to properly teach him or her.

Summary

The category *Assessment* is vital for learner and teacher or researcher alike. Both the teacher or researcher as well as the learner need to know how the learner is doing. The teacher or researcher wants to know whether the game is effective and the learner is progressing, and the learner wants to know how close he or she is to achieving the goal (as well as to learn from mistakes). The designer should know what the eventual real-life context requires from the learner to make sure that the right skills are trained.

Conflict / Challenge

Prior knowledge

In educational games, a player is challenged on multiple levels:

1. one has to know how to interact with the system and know how to achieve goals through the controls of the game; and
2. one has to have a sufficient grasp of the educational content to be able to overcome the domain-related challenges presented by the game.

Both of these types of knowledge have to be kept track of to make sure that the player receives the optimal challenge. Not taking game play knowledge into account can lead to the navigation problems described in the introduction (Tawfik, Moore, He, & Vo, 2012). Not teaching any educational content before presenting challenges can bring frustration, boredom, or trial-and-error behaviour. The latter can be seen in Ke's (2008) article on small math games aimed at 4th and 5th graders, where a lack of domain-knowledge prevented the students from learning without further assistance.

When confronted with new information, we initially try to interpret this information within existing knowledge schemas (Hattie, & Yates, 2014). By doing so, we give more relevance to what is to be learned, i.e. we embed it in what we already know. Another benefit of this process is that it supports recall at a later moment; the more connections we can make to existing knowledge, the easier it is to remember the information (Tulving, & Donaldson, 1972). Thus, determining the player's prior knowledge and adapting the detail or complexity of the instructions to it can greatly help in not only teaching the player, but also engaging him or her, as the information is seen as more relevant.

Transfer: Memory Demands

Ideally, the skills a designer wants to assess determine the type of conflict presented to the player. If one wants to know whether the player has learnt how to calculate a derivative, one presents the player with a challenge which requires a derivative to be calculated. Challenging the player to calculate a primitive instead would not tell the designer anything about the player's skill in derivative calculations (and would be interpreted as unfair by the player).

Although this example seems silly, the point it makes is valid: let in-game conflicts resemble the real-world. This does not automatically mean they should be exactly the same, just that the same is asked of the player. The difference between the *Performance Change* section described earlier is that keeping the scoring the same is only a part of the challenge, the type of conflict is the other.

Problem clarity

While related to the section on prior knowledge, the measure of structure a game provides for its conflicts also determines the challenge. The less structure a conflict has, the more the learner has to know upfront about such conflicts (Rowe et al., 2010; Hattie & Yates, 2014). This can be balanced by providing structured conflicts early on and gradually make each conflict more abstract as the player progresses (van Merriënboer & de Croock, 2002).

Greitzer et al. (2007) also mention proper scaffolding as a key element for good educational game design:

"Devise learning scenarios that maintain the performance of learners in a "narrow zone" between too easy and too difficult. This ensures the player is challenged by the scenarios, but not intimidated; helping the learner to gain self-efficacy and to remain engaged in the scenarios."

Summary

Educational games require instructional design to guide its players on two fronts: game play knowledge and domain-specific knowledge. Adapting the game's complexity to the player's level of both types of knowledge can help the player learn the game's educational content effectively. Of importance here is that the game challenges the player with conflicts that resemble the conflicts' real-world counterparts.

Control

Active learning

The category *Control* refers to the amount of agency and optional ways of interaction the player has in a game. The more closely a player can determine their own actions and methods of approach, the more actively they are involved with the game play and in extension the educational content (Garris et al., 2002; Greitzer et al., 2007).

However, giving the player more agency also means that he or she has more *Action Language* to be aware of in terms of interface and options. Executed poorly it can gravely distract the player from the educational content. For example, a 3D environment that needs to be navigated by the player gives a lot of agency to the player as they get to decide where to go. When not implemented properly, this can cause the player to just wander about, not interacting with the educational content and instead simply enjoying the freedom.

Giving the player more agency also does not imply that the game can let the player figure things out on their own. Properly guiding the player through the learning experience using clear goals, instruction, and feedback is found to be more effective than letting the player discover and learn completely on their own (Clark et al., 2010).

Representational Congruence

An important aspect in choosing the options and agency of a player is looking at the real world. Transfer is more easily achieved when the options the player has are closer to the ones they will have in the real world (Holbert and Wilensky, 2012). Representing this as closely as possible is referred to as making the game *representationally congruent*.

Holbert and Wilensky's (2012) game FormulaT successfully transferred the learned knowledge to real world problems. In the game, the player has to pre-plan a racer's speed by manipulating a velocity over time graph of the parkour. By using the same graphs as the players would have to reason with in the real world the designers gave the players a lot of freedom, but with the same tools as they would use in the real world.

Summary

While giving a player a lot of agency makes the game a more active learning experience, it is the task of the designer to make sure that the player's freedom is focused on interacting with the learning content. This can be achieved by providing the player with the options they would otherwise have in the real world.

Environment

Transfer: Physical Dimension

The category *Environment* describes the digital world the player is operating in. Unrealistic settings such as sci-fi worlds would mean a large distance between it and the real world. Results of games with far wilder environments show that this is not a requirement for non-simulating educational games (e.g. barab et al, 2009). In terms of transfer, it is much more important that the learned skills are not too instantiated in the world (Bransford & Brown, 2000). This concerns simple things, such as using the correct terms and words, but also adhering to general rules found in the real world (e.g. laws of physics).

Environments are more often used for their engagement appeal, as a sci-fi or medieval setting is more interesting for the player than the real world setting. The distance this would create can be mitigated somewhat by making sure that the fantasy is relevant to the educational content (Habgood et al., 2005)

Summary

While the appeal of games often comes from fantastical environments not seen in real life, the underlying logic of the world and the representation of the educational content should not suffer from this. Instead, the choice in environment should be relevant to the to-be-taught knowledge.

Game Fiction

Relevance

Having the player progress through a story-line can make a game more engaging, immersing the player into a world in which they play an integral role (Garris et al., 2002). However, research on the learning gains with and without a narrative structure indicate that using a narrative can distract the learner from the learning material (McQuiggan et al., 2008). This makes sense, cognitively speaking, as the player is not only interpreting the game play and educational content, but also engaged by the story. The player is thus paying a lot of attention to the story, as it is a very engaging part of the game.

Having the story structure be relevant to the educational content can mitigate this problem to some degree (Dickey, 2011). In the game *Murder on Grim Isle*, the player needs to solve a mysterious murder case. As the goal of the game is to teach argumentation, the story closely fits the learning content, and as a result the learned skills transferred to a writing assignment.

Summary

Game Fiction is a very important category for creating an engaging game, but the learning goals of the game have to be taken into account when writing the story. Storylines irrelevant to the educational content will only distract the player from thinking about what he or she is learning.

Human Interaction

Cooperative or competitive

In terms of interaction types and their effects on cognitive processes, there is little research to be found in the field of game design. Research that has looked into these different kinds of interaction between learners often looked into the responsibility structures (i.e. do students work together or is there competition involved). Similar to non-game related research (e.g. Johnson et al., 1981) educational games also benefit more from a cooperative working structure than an individualistic one (e.g. Ke and Grabowski, 2007). However, there are mixed results when it comes to discerning differences between cooperative and competitive (Plass et al., 2013).

Transfer: Social Dimension

Reasoning along the line of the social dimension, transfer would occur more easily if the social structure in which a skill is learnt resembles the social structure in which the skill should transfer to. However, even in literature on transfer there is little known on the influence of learning a skill in a group instead of on one's own (Druckman and Bjork, 1994).

Summary

For our intents and purposes, this category needs more research before we can really discern useful cognitive elements.

Immersion

Immersion and Deep Learning

The category *Immersion* describes the player's view of the digital world and story of the game. The more the player is immersed, the more willing the player is to accept the rules and knowledge of the game (Garris et al., 2002). This can lead to an interesting problem in which the player is so immersed in the digital world, that the knowledge gained in that context is not going to transfer anywhere (Habgood and Ainsworth, 2011).

Barring transfer, immersion reflects the engagement of the learner with the game and thus indicates more engagement with the educational content. Thus, an immersed player will learn more, and effort should be taken to maintain immersion whilst providing an individualized learning experience (Kickmeier-Rust and Albert, 2010).

Intrinsic Integration

A game mechanic can be described as a method invoked by agents (players) for interacting with the game world (Sicart, 2008). It is the translation of the *Action Language* to determine the amount of *Control* the player has. A game's mechanics represent the in-game actions the player takes to achieve the in-game goals.

Part of creating an immersive experience and making it a strong learning experience at the same time is making sure that the learning content is integrated in the game's mechanics. An example

Figure 2.3: A screenshot of the game Zombie Division (Ainsworth, 2013). The number '32' on the chest of the skeleton can be divided by four, which the player has to select to defeat the skeleton.

of this is seen in *Zombie Division*, an educational game designed to teach mathematical division (see Figure 2.3). Habgood and Ainsworth's (2011) game revolves around laying skeletons to rest by selecting the right weapon to slay them. To know which weapon is the correct one, the player needs to identify the number on the skeleton's chest and select an available weapon by which the skeleton's number can be divided by. Each weapon has a corresponding number, ranging from two to ten.

In this example, the game mechanic used most often (the method of fighting) is mapped onto the educational content. The difference between *representational congruence* and *intrinsic integration* is the fidelity in which the educational content is represented. In the FormulaT game, making players use actual velocity over time graphs results in a much stronger resemblance to the real world application of the trained skill.

Neither form is necessarily better, it is the strength of making the main interaction with the game involve the learning content which makes it an effective design principle.

Summary

Having the player be immersed in a game is often a good sign of player engagement. For an educational game designer it is important that the engagement with the game equals engagement with the learning material. This makes sure that the player is immersed in the learning content with the educational game being the medium.

Rules / Goals

Clarity

Rules and *Goals* are at the heart of any game, be it digital or analog. Having clear goals presented to the player helps to provide a clear focus and, in turn, guide the player towards a more engaging game play experience (Greitzer et al., 2007).

Relevance

As established throughout the other categories, they need to be relevant for the educational content. *Goals* and *Rules* are no different, as the player will want to achieve the game's goals,

these will naturally have to align with what the underlying educational goals are (Shelton and Scoresby, 2011). If this is not done correctly the player will be focused on achieving a goal which has nothing to do with the educational content, hindering the learning process as attention is focused elsewhere (Squire et al., 2004).

Working Memory

As the player is learning educational content whilst progressing through a game, it is vital to help him or her by providing knowledge such as rules at the proper time. This keeps the amount of information that has to be remembered low, allowing the player to learn the rules slowly (Bellotti et al., 2010). Pop-ups, highlights and other visual or audio instructional methods can be used to remind the player about goals or rules.

Summary

The rules and goals of the game provide the player with a reason to play and a ruleset on how to do so. The player's mental model of the rules grows as the game progresses, and as such, the player can be expected to know the rules by heart. Early on, however, the player will need to be reminded of the rules and goals of the game to help him or her learn all the new information in the game. It is essential that the rules and goals of the game are relevant to the desired learning outcomes, as otherwise the player is learning irrelevant information for an irrelevant goal.

Summary

This chapter started with the question:

"What is the relationship between the elements of an educational game and the cognition of the user and how do they relate?"

For this thesis, the answer is found mostly in the game elements related to the player's input, the game's output, and how the player's working memory is supported by the game. Some key points found in this chapter:

- *Relevance:* ideally, all categories of the game should be relevant to the learning content to make sure that the player's engagement with the game leads to achieving the learning goals.
- *Guidance:* several categories (e.g. Conflict / Challenge, Control, Rules / Goals) have a large impact on what is expected of the player and by extension whether the player can cope with all the new information. The interaction of these categories and the resulting complexity of the game should be kept graspable for the player.

Now that we've identified the relationship between the elements of an educational game and the cognition of the user, it is clearer where and how cognition can play a role during the design process. The next chapter will introduce a problem domain in which our assumptions will be tested.

3. Problem Instantiation

To properly understand the influence of our design assumptions (i.e. including cognition into the design process and using non-game related design methods), we need to create a game in which to test these assumptions. For that, we need a skill to train. A knowledge domain should provide a set of complex skills for which an educational game would be a good training tool. The conclusions based on the results of the tests in this domain will however be focused on the applicability and effect of the non-game related methods, and not necessarily in results concerning the domain.

Domain and problem

The domain will be the ergonomic safety of professional caretakers in nursing homes, hospitals and other institutes. The physical and mental stress Dutch caretakers face in their working environment is causing burn-outs, long-term physical pain, and other problems in a large population of this workforce (Bronkhorst, Ten Arve, Spoek, & Wieman, 2014).

One of the problems causing the physical strain on caretakers is the working routine of the caretakers themselves. Proper posture and working ethos is taught barely in nursing schools, and caretakers quickly develop their own working behaviour at the start of their career. Sadly, this causes the proper techniques to fade away over the years as they are replaced by methods that seem harmless, but are devastating for caretakers' physical health in the long run.

While this problem has multiple aspects and is not easily solved, the skills involved in nursing provide a good basis to apply the literature to. How does one properly handle the scenarios which take place in a nursing home? For example, how do I lift a patient from his or her bed to a wheelchair without endangering both the patient's health and my own? The potential complexity of this problem increases when contextual factors such as materials, space, and patient characteristics are introduced.

From an educational point of view, teaching a skillset flexible enough to deal with these differing contextual factors is the main problem. Simply teaching all possible scenarios is an unrealistic approach, as it is difficult to create such an exhaustive list let alone remember it all. We believe that providing caretakers with a more expert-based approach consisting of recognizing similar problems and having general guidelines on how to deal with such similar problems is a more realistic and desirable solution.

This leads to the question: 'How can one teach a complex skill?'

Teaching a strategy

For this thesis we wish to teach only a part of this flexible skillset: the skill to solve a patient-repositioning problem. During the course of a day, patients with severely impaired movement have to be repositioned from their bed and wheelchairs to toilets and showers. These logistical problems often have multiple angles of approach and complications caused by the patient's room, other caretakers and patients, and the willingness of the patient him- or herself. These factors can differ wildly due to the nature of the job: working with patients. For example, rooms often have different interior designs, the types and brands of aiding-tools can differ between patients, and the patients themselves often require a specific approach. This is only a set of the factors that influence the complexity of the repositioning problem.

To maintain the health of the caretaker as well as the patient, the caretaker needs to be mindful of the options he or she has and when to apply which option. This comes down to *strategic* knowledge, as the caretaker has to know not only how to execute an approach, but also why that approach is the correct one for the encountered problem.

Burns, O'Donnell, and Artman (2010) have shown that a high-fidelity simulation can help teach the nursing process that nurses learn in their first year. This is a problem-solving process and, while our repositioning problem is less complex, shows similarities in how caretakers should think about moving a patient.

The caretaker has to assess the variables that play a role in determining the solution, diagnose the right approach, plan the availability of all tools and prepare the patient, implement the chosen approach, and finally determine whether the chosen approach was ultimately correct or whether a different approach would have been safer.

As such, we believe that an educational game that resembles the repositioning problems caretakers encounter during their work can provide the learning experience needed to learn a useful strategy for determining the right angle of approach in solving such a problem. This strategy will focus mostly on the assessment, diagnosis, and evaluation aspects of the process, as these lend themselves well to a digital setting, whereas the implementation and planning phases would be hard to emulate on a computer as these would not match their physical nature in the real world. Also, this complex skill requires a lot of decisions to be made, making it cognitively demanding and thus interesting for our user-cognition sensitive design approach.

Research questions

In this thesis we want to explore the application of non-game related methods as a means to formalize the design process of an educational game and more fully embed cognition into the design process as a whole. The problem used to test these methods on is situated in the healthcare domain and deals with training a domain-specific problem-solving skill.

Questions on embedding cognition

As discussed in the background literature, *feedback* is one of the strongest moments for learning. That is, however, provided that the feedback type and timing are fitting for the kind of learning moment one wants to have. The skill defined in this chapter focuses on the analysis of a situation and the identification of critical factors in that situation which determine the correct repositioning approach. Not properly analysing the factors would lead to errors in the chosen approach.

Literature suggests that reflective feedback stimulates the reflective and critical thinking required to properly analyse the situation (Wetzstein & Hacker, 2004; Zohar & David, 2008). Thus an interesting test case would be to see whether reflective feedback leads to a better learning experience than a type of feedback that does not take the required critical thinking into account. The feedback for the test group will analyse the error of the user to identify which contextual factor was not taken into account and then ask the player whether he or she has considered that factor. The other group will only receive feedback on what part of their approach was wrong and what that part should have been. This leads to our first research question:

1. *Do participants who receive reflective formative feedback on their errors perform better at analysing the situation than participants who receive corrective feedback?*

Another aspect of cognition that was identified in the background literature was the effects of *prior knowledge*. The design of our game will take into account the possible lack of prior game play knowledge, but it would also be interesting to see if there are any major performance differences between participants with gaming experience and participants without gaming experience. This leads to the second research question:

2. *Do participants with experience in playing games perform better at analysing the situation than participants who have no experience?*

Question on effectiveness structural approach

A different matter is the structural approach we will take in designing the educational game. We want to know whether using this process still leads to an effective game, and ideally how our approach contributed to the effectiveness. For the effectiveness of our game, we want to answer a third research question:

3. *Do participants become better at analysing the situation throughout playing the game, and if so, does the test group learn better than the control group?*

Our last research question remains more open, as the scope of this research project is too small to empirically compare complete design methods. As formal non-game related methods have hardly been used in educational game design, the following experimental research question is posed:

4. *How can we structure the educational game design process using non-game related design methods?*

Sharpening the questions

Please note that the first three research questions will likely change depending on our game's design. There has to be a formal way of measuring how good a player is in analysing the situation. The research questions will be revised appropriately when a suitable method has been designed.

4. Preliminary Research

Now that a domain has been established to apply the knowledge to, we need to answer several questions concerning this domain:

1. How do nurses solve a patient-repositioning problem?
2. What kind of mistakes do they make?
3. How are nurses trained on solving these problems?
4. Are there general guidelines, and if so how are they enforced?

As discussed in the previous chapter, a lot of these questions have a cognitive aspect. The patient-repositioning problem has some inherent decision making processes which involve cognitive, conscious, choices. Therefore, we believe a cognitive task analysis (CTA) can help us in answering the above questions.

Cognitive task analysis

A CTA method is always applied to a representative individual of the domain one wants to analyse (Clark et al., 2008). These *Subject-matter Experts* (SMEs) can be interviewed and observed in an effort to acquire a greater understanding of how a task in their domain of expertise should be executed. The knowledge extracted from these SMEs can subsequently be used to design a training method for teaching the extracted knowledge. How we chose to design our training is discussed after the CTA section.

Aside from the decision-making process, we needed to know whether the problem of incorrect posture found in Bronkhorst et al.'s (2014) report were due to a lack of understanding the consequences of one's behaviour, or whether there were other, yet unknown, reasons.

CTA method

Participants

During January and February 2016, four interviews were carried out. Two (Dutch) SME's in the field of healthcare, proper execution of tasks, and safe work environments for nurses were each interviewed twice.

Materials

Two structured interviews were conducted with each expert, the first focused on formalising the thought processes involved throughout the execution of caretaker tasks, and the second focused on identifying possible pitfalls and other reasons for why caretakers, albeit unconsciously, endanger their own health. In the first interview we provided example problems which had to be resolved properly and the interviewee had to explain his or her reasoning behind each decision. For the second interview, two different techniques were used:

- card-sort tasks of the official task classification. One for physical strain, endangering the patient, and decisions being context dependent (each a separate card-sort task); and
- a free-recall on possible pitfalls and other reasons for why (and how) nurses endanger their own health, per official task class. The recall was semi-structured by providing three categories to classify their reasons with; knowledge related, behaviour related, and logistics related.

The reasoning behind this two-way structure was to go more into detail during the second round of interviews using the information gained from the preceding interviews. The card-sort task allowed us to use the official task classification as taught in nursing schools to determine which

class of tasks requires the most decisions and assessments. The official task classification is as follows:

Category 1: Transfers within the boundaries of the bed.[6]

Tasks from this category include: turning the patient onto his or her side, helping the patient sit at the edge of the bed, etc.

Category 2: Transfers from the bed to and from (wheel) chairs and toilets.

The choice in tools used for tasks in this category depends greatly on the mobility of the patient. From simple assisting tools for mobile patients to electricity-powered hoists for patients who can no longer walk, the nurse needs to know when, and how, to apply what tool.

Category 3: Putting compression stockings on or taking them off.

Although this category's tasks do not require explanation, it is important to note that there is a plethora of assisting tools the nurse can encounter during his or her career. It is important to know the small distinctions between each brand and kind for the nurse to put them to good use.

Category 4: Static load and working in difficult positions.

This category is not comprised of tasks, but focuses on the more invisible dangers such as the always present strain caused simply by standing, walking, lifting, etc. It also includes guidelines to maintain a proper working posture.

Category 5: Manoeuvring rolling materials.

Tasks from this category include: the movement of wheelchairs, beds, and other rolling materials such as hoists.

Results of the first round of interviews

The most striking general finding was the amount of caveats and different constraints that could influence how a task has to be executed. Although the general description of each task (as shown in official instructional videos provided by one of the experts) gives a clear overview of the steps of a task, these descriptions lack realism.

For example, some tasks require a room in which the patient's bed can be moved around freely in order for the nurse to be able to maintain a proper posture. However, both experts stated that being able to move the bed often isn't possible in the real world. In such cases, if the nurse wants to maintain a healthy working posture, a broader set of possible correct solutions has to be known.

Other problems such as the emotional and mental state of the patient or tubes and other wires hanging around the bed can increase the complexity of a task considerably.

More specific results showed that the second category requires the most decisions and planning, suggesting it is the most complex set of tasks.

Results of the second round of interviews

The general finding of the card-sort task revealed something which the preceding interviews had partially suggested; categories one and two have the most demanding tasks.

[6] A transfer task aims to reposition or move the patient.

The specific findings per sortation-category were as follows:

Physical strain

Both experts put categories **one**, **two**, and **four** as their top three most physically straining. Categories one and two contain not only the most physically demanding tasks, but are also the most prevalent of all the tasks.

Category four comes in third partly due to the prevalence of categories one and two, as doing the same tasks a lot during one work shift is even more straining. Aside from frequency, category four also influences how physically straining the rest of the tasks are, as a bad posture increases the negative impact of other tasks on the nurses health.

Category **three** is not as straining due to the assisting tools available. Category **five** is the least straining, as the task of moving rolling material around is well defined, making it much easier to perform those tasks.

Endangering the patient

Both experts placed category **two** as most dangerous, number one reason being a wrongful estimation of the mobility of the patient, which is also influenced by the emotional and mental state of the patient. The order of the rest of the categories was very similar for both experts again, the reasoning behind the ranking coming down to nurses being too hasty and harsh.

Categories one and five take the second and third place (albeit in a different order per expert), followed by the third category. Both experts stated that category four does not endanger the patient.

Context dependent

Categories **one** and **two** are both placed as number one here, as they depend a lot on the space the nurse has to perform the task and the state of the patient (both mobility-wise and mental). The other categories require less context-dependent decisions to be made and are thus ranked lower.

Open-recall task

The general finding of the open-recall task was that there is a diversity of behavioural and contextual causes for the fact that nurses are not maintaining a proper working posture. It made the initial idea for a purely educational intervention seem insufficient as a catch-all answer. The behavioural causes can be summed up in the following three aspects:

- Caring behaviour leading to neglecting nurses own health.

Bronkhorst et al.'s (2014) report states that nurses are not incentivized to take care of themselves by the institutes they are working at. Both experts stated that nurses also feel a sense of responsibility. This sense of responsibility culminates itself by trying to keep the patient from helping the nurse during a task. In a sense, nurses feel like they should bear with it, as it is their job. This also goes for problems that fall under the patient's (and the family's) responsibility, such as keeping the patient's room spacious enough for a nurse to walk around the bed.

- Wrong working etiquette as a result of erroneous education and mimicking.

The right way to perform the tasks is not taught enough during a nurse's education nor sustained during the switch from school to working environment. Interestingly, both experts stated that learning on the job is often counterproductive for the working posture of the new nurse. Most experienced nurses have developed their own (often wrong) working style, and their confident

and fluent performance leads new nurses to copy the behaviour (keeping the wrong behaviour from fading away as older nurses retire).

- Negative repercussions of bad working posture culminate mostly over the long-term, impeding a reality-check for nurses and preventing change.

Doing the same physically straining actions too often per day eventually leads to physical injuries. An example of this is RSI due to a wrong working posture as seen in desk jobs. However, these injuries are not apparent instantly after performing a task with a bad posture, instead it builds up over time and symptoms slowly introduce themselves.

This reason was acknowledged by both experts to impede change in the nurses working posture. Performing tasks in a non-routine way costs more time and concentration, making the job initially more demanding of the nurses as the proper routine has to become their own. Add to this the fact that nurses do not experience any negative consequences of their usual routine, and a situation arises in which change is difficult to achieve with small interventions (e.g. a yearly two-day workshop). The same goes for the mimicking nurses; they only see the benefits as any injuries the experienced nurse has are not visible either.

Other results showed that nurses often lack more detailed knowledge of aiding tools, making it more time-consuming to use an electricity-powered hoist for example. This causes nurses to perform tasks manually (often at the cost of one's correct working posture), or call in the help of another nurse even though the aiding tools are meant to substitute an extra nurse.

Conclusion CTA

The diversity of problems nurses have to solve highlights a need for a mental toolbox suited to perform the same task in a multitude of ways. This toolbox has to be accompanied by an understanding of when to use what tool. Thus the CTA not only provided a more thorough understanding of the tasks a nurse has to do, but it also deepened our understanding of the problem. This more detailed description of the problem in turn helps us to create a better solution for the problem. That solution being: providing a 'toolbox' for the nurses and the knowledge on how to use it.

Teaching the entire toolbox goes beyond the scope of this thesis, as it ideally would involve multiple kinds of games combined with some form of real-life coaching (to train proper execution). Instead, as mentioned in the previous chapter, we wish to focus on one *part* of using the toolbox: being able to properly analyse a situation and use the gained information. This allows us to engage our users in a realistic problem-solving challenge, something which is also found to be educationally effective (Greitzer et al., 2007).

As seen in the CTA, there are a lot of factors involved in determining the right approach to a situation. Thus we need to teach caretakers to think more about these factors and their consequences for the caretaker's chosen approach. The chosen skill is complex due to the decision-making processes involved and the hazardous contextual factors that have to be *recognized* as such. In the next chapter, we discuss a prototype which we will expand upon and determine what kind of non-game related method we will use to design the additions and changes.

5. Method: Design

This chapter describes the design phase prior to the experiment. It will cover the prototype we used as a starting point, followed by the description and application of the 4C-ID method. Finally, the resulting changes that were implemented as a result of the 4C-ID method and the acquired knowledge of the CTA will be discussed.

The game: 'How to help your dragon'

The prototype of 'How to help your dragon' (HHD) was designed by first year students of the Game Design and Development Major at the Hanze Universiy of Applied Sciences to teach caretakers the guidelines on what the human body can physically handle without being overtaxed. The setting provides a problem similar to the real world, safely helping a person (or in this case an egg) from one place to another. As discussed in the background literature section, a fantasy setting can make it harder for players to identify the analogy with the real world, making it harder for transfer to occur. Thus the game's fantasy setting will limit transfer and pose a limitation on the amount of realism we can incorporate in the game's mechanics.

Prototype game play

In HHD, the player is expected to create a set of actions for the in-game avatar to execute (see Figure 5.1 for an overview of creating an action sequence). The goal of these actions is to make the avatar move a dragon egg from his nest to a campfire in order to help the egg hatch.

Step 1: Select an empty slot at the bottom of the screen to reveal the two different categories of actions.

Step 2: Select one of the categories to reveal the possible actions of that category. Clicking the trash bin will remove a previously selected action.

Step 3: Click on the action you want to put into that specific slot. Clicking on a question mark next to an action will provide a pop-up with more information about that action.

Step 4: The selected action is now visible in the sequence. One can edit the action by clicking on it and going through steps one to three again.

Figure 5.1: An step-by-step explanation of the action sequence creation procedure. The action bar has eight slots, but it is not required to use all of them.

The player can choose actions from two categories:

1. Manual, actions such as lifting, carrying, pushing, etc.;
2. Instrumental, invoking tools such as wheelchairs, hoists, and medical gloves.

The right solution depends on the weight and temperature of the egg, as it determines whether the avatar can lift and carry the egg, or whether an additional tool is required. Constructing the wrong set of actions will lead to summative feedback at the first erroneous action, showing a red overlay over the action and having the avatar reach for his back in agony. After the feedback the player regains control and can edit the chosen actions to create a new plan, allowing him or her to approach the problem with a set of hypotheses. Note here that the feedback does not explain anything, so the player will often have to identify the nature of a mistake on a trial and error basis.

Small step-by-step explanation of one level

The single level designed by the Hanze students was made to showcase the game play. To provide a better understanding of the game play, we will provide a step-by-step explanation of how this level is played.

Before the level starts, the player is presented with a small introduction screen displaying the goal of the game and the relevant contextual factors (weight and temperature of the egg). The weight of the egg is 20 kilograms, the temperature is 70 °C.

The egg is too heavy to carry it all the way to the other side manually, so a tool will be necessary. The egg can be lifted manually, so a wheelchair is chosen. As this means that the egg has to be touched by the character, heat-resistant gloves have to be put on. The subsequent movement actions and resulting action sequence can be seen in Figure 5.2.

Finally, the player presses the ‘play’ button in the corner to execute the action sequence, and the egg reaches the other side safely.

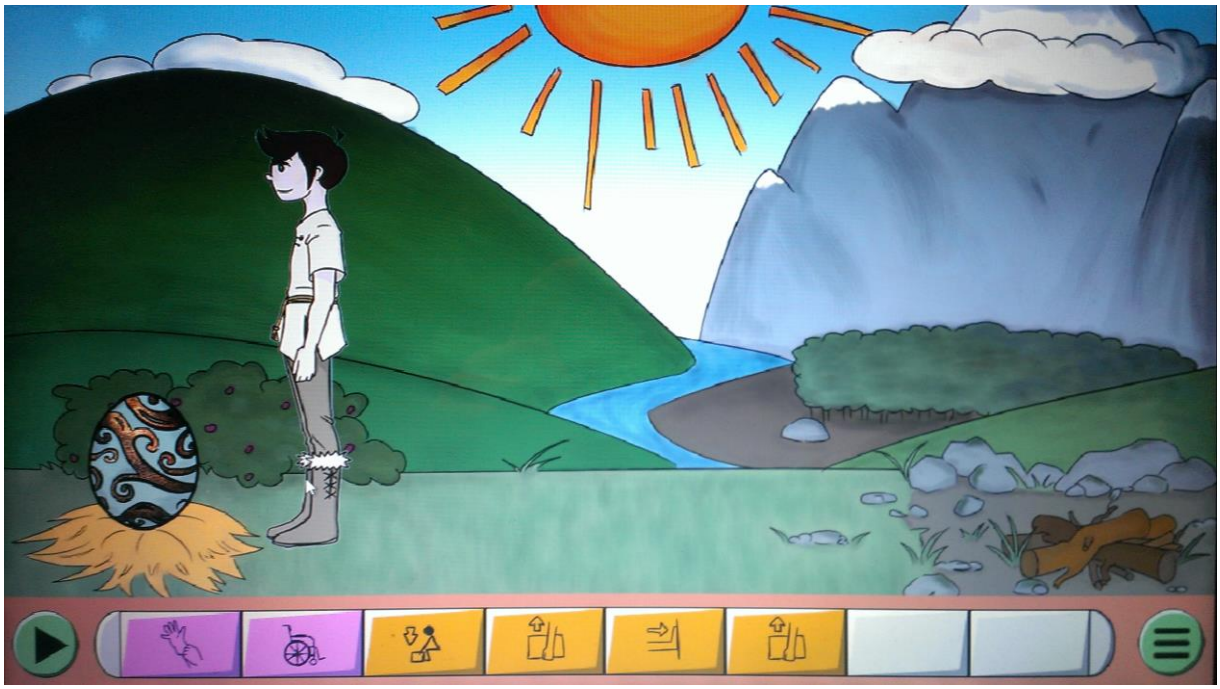


Figure 5.2: An screenshot of the only level of the prototype. Note that not all slots have to be used for a solution.

Game's applicability to our problem

HHD's strength lies in the fact that the player has to think of an approach, test it, reflect on the feedback, and then try a new approach in case the first one was wrong. This cycle bears a strong resemblance to the Nursing Process cycle discussed earlier, making it an interesting base for our learning goals.

Intended changes based on CTA results

Based on the CTA results and the state of the prototype, we have the following changes:

1. Feedback: the prototype only provides summative feedback which does not support the learner in understanding their errors. As discussed earlier, we want to test a reflective and a corrective type of feedback, both of which have to be implemented.
2. Amount of levels: the prototype consisted of only one level, which is too little to learn the skill we wish to teach.
3. Prominence of contextual factors: as is, the prototype only has two contextual factors, both of which are given at the start. Doing so makes it no longer necessary for the player to analyse the situation him- or herself. Our game requires this analysis phase to be a more prominent aspect of the problem, thus some additional restriction has to be designed to require the player to analyse the situation.
4. Amount of contextual factors: aside from the lack of importance of contextual factors, the two there were now left little possibilities to design with. Additional factors have to be designed to allow for more varied levels.

Teaching a complex skill: 4C-ID

Our goal is to teach a complex skill: assessing a repositioning problem, deciding upon the best approach, and then implementing said approach (where we are most interested in the first two steps). Four component instructional design was created with the idea of teaching a complex (cognitive) skill in mind and thus suits our instructional needs (van Merriënboer, Jelsma, and Paas, 1992).

The four components

The 4C-ID consists of four interrelated components, all playing their own part in the learning process. These parts are:

1: Learning tasks

The basic structure of the 4C-ID method comes from a sequence of concrete, authentic, whole-task experiences. As described in the previous chapter, whole-task, realistic, experiences are recommended for immersing the learner in the problems they are solving by giving them relevance through context.

In 4C-ID, a sequence of learning tasks is defined as a *class* with the intention of gradually adding complexity as one proceeds from one class to the next. Each class has varied whole-tasks of similar complexity, requiring the same body of knowledge to be completed.

Scaffolding is applied to make sure that players are supported strongly at the beginning of a task class, as they require the most aid during the first tasks of a class. The amount of support decreases as the learner progresses through the tasks of one class. This helps in keeping the cognitive load low when the learner is still grappling with the new information he or she is presented with.

In Figure 5.3 the task classes are represented by the sets of large circles denominated by the dashed lines. The circles represent task classes, with the triangle representing diverse scenarios and the grey fill representing the scaffolding (which decreases per task class).

2: Supportive information

As with any instructional method, some form of declarative knowledge has to be taught for the learner to be able to learn a complex skill. Supportive information is given before the start of each task class to bridge the gap between the learners existing knowledge and the knowledge needed to finish the learning tasks of the task class. Subsequent task classes are preceded by supportive information that only elaborates on existing parts of the created mental model or add additional knowledge to it.

Supportive information is geared towards preparing the learner to solve the non-recurrent aspects of a task, i.e. it provides the general know-how for approaching a task and solving it, but does not provide training in the procedures themselves.

In Figure 5.3, the supportive information is shown as the dark coloured blocks before and under each task class (making somewhat of an 'L' shape).

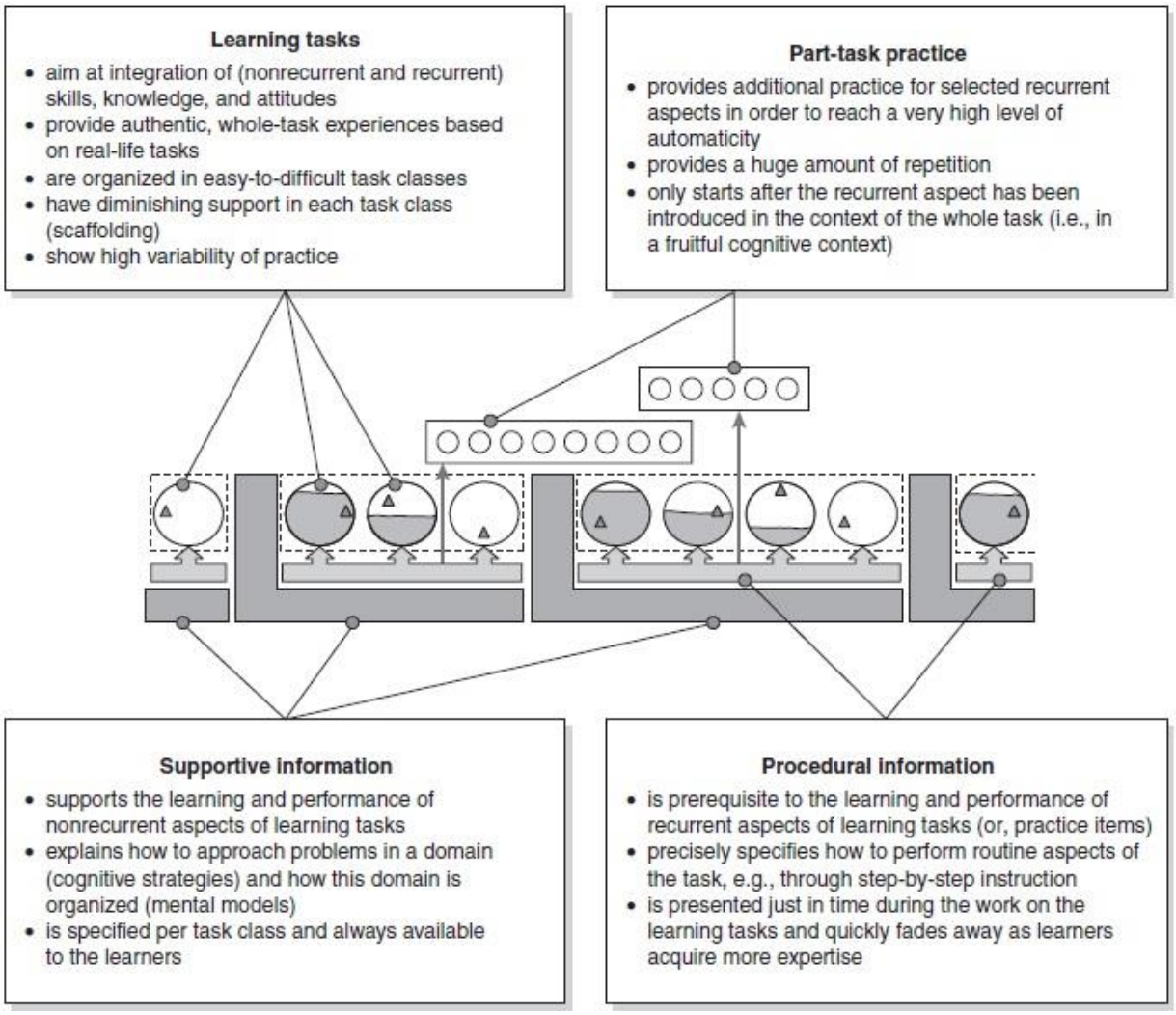


Figure 5.3: An overview of the four core components of the 4C-ID method. The triangle in each learning task (i.e. the circles) indicates the variety between the learning tasks in one class. The fill in each learning task resembles the amount of scaffolding provided to the learner. Image taken from Kirschner and Merriënboer, 2008.

3: Just-In-Time information

Just-In-Time (JIT) information is there to train the recurrent aspects of the learning task. It reinforces the rules and procedures of which the complex skill consists *during* the learning tasks themselves. This is done through feedback, as well as information displays the learner can call upon from the interface.

In Figure 5.3, JIT information is shown as the thin bar underneath each task class, its interaction with each task in the class is represented by the arrows.

4: Part-Task Practice

A particularly large complex skill (including a lot of constituent skills) can require additional training of smaller parts of the task to acquire the required fluidity for said part of the skill. Especially when the next, more complex, task class requires a certain level of mastery of specific parts of the whole skill, the fluidity gained from such part-task practice helps reduce the cognitive load imposed by the increase in complexity. JIT information is also useful for this component, as it also focuses on training a recurrent aspect, only of a part of the complex skill.

In Figure 5.3, these part-task practice sessions are shown as bars with smaller circles in them, the circles in such a bar represent the small tasks to train a constituent skill.

Together, these four components form the base of the instructional design method called 4C-ID (see figure 5.3 for an overview of the resulting instructional design structure).

Instructional design using CTA results

In our case, we use the four components as a structuring tool for our CTA results. The resulting structure will then be used as an instructional blueprint for the actual educational game design. This allows us to design the additions and changes to the prototype without the fear of straying from the instructional design.

We will discuss each of the four components in terms of what they contribute to our instructional design:

1: Learning tasks

The task classes are there to: 1) familiarize the learner with the game's interface and game play; 2) present challenges; and 3) test the learner whether he or she has gained enough understanding of the skill to handle more difficult challenges. Each of these three goals have their own task class with corresponding complexity, ranging from goal one being the first and least complex class to goal three being the last and most complex class (with goal two being the second class with average complexity).

The learning tasks of these classes will have to provide a variety of ways in which the skill has to be applied, also requiring different tools to promote familiarization with each of them and stimulate the learner to think about what tool to use for a given situation.

2: Supportive information

In terms of supportive information, we need to make sure that the player is aware of the game's rules and controls before the first task class. The goal and contextual factors of each learning task have to be provided before each task, but should not require memorization (to prevent cognitive overload).

At the end of each learning task, the player will receive more general feedback on their performance as a whole during that task in addition to the feedback provided during the task. The distinction between these two is the specificity, with the JIT feedback focusing more on the error and the supportive feedback focusing on the general strategy.

3: Just-In-Time information

Feedback should be provided after errors to give some JIT information on what the player should have done or should think about in their next attempt.

The selectable actions and tools already have question marks added in the interface with which the player can interact to learn more about an action or tool. Contextual factors should either be clearly visible, or easily retrievable, from the interface.

Finally, during the first task class, the player should be able to familiarize themselves with the game play more easily by being able to reference back to the supportive information provided at the start of that task class.

4: Part-Task Practice

Of the aspects this game wishes to teach, one could definitely add part-task training for the diagnosis-aspect of the game in which the player determines an approach using the contextual factors at play. For the scope of this thesis however, we have chosen to refrain from doing so.

Resulting instructional structure

Bringing the components together leads to the task class structure as seen in Figure 5.4. The first three tasks constitute a tutorial in which the player is familiarised with the game play. The second task class consists of three smaller classes of difficulty, each with three tasks of their own. The last task class adds a new factor to measure, distinguishing it from the second task class in terms of required knowledge and complexity.

The actual mechanics were not designed through using the 4C-ID method however. The method only gave requirements for, for example, the structuring of difficulty, instructional guidance, and when feedback should be presented.

Conclusion

The methodical way of the 4C-ID method requires its user to really think of the interconnections between each of its components and how one want to implement them. The task class structure translates very easily into level design when concerning its application to game design.

However, a lot of the examples used in the teaching aspect of 4C-ID relate more easily to classic instructional design than educational game design, such as using worked examples, inquiry based teaching, and more.

Educational game design

Using the general structure provided by the 4C-ID method as a whole, the more detailed blueprints provided by our instructional design, and the knowledge gained from the CTA results, this section will describe the reasoning behind the additions and changes. Note that these additions and changes, while built on the knowledge and requirements gained from the two non-game related methods, were still made into game components and mechanics through brainstorming. However, most of these game components are almost direct translations from scenarios discussed with the SME's.

Figure 5.4: An overview of the task class structure. Note that the final free tasks are not actual test trials, but provide a more complex challenge for the player which allows us to see whether they can cope with a new challenge.

Measuring tools

In the initial prototype, the player had access to all contextual factors right away. This lead to the problem of having perfect information, resulting in the fact that the player needed little reasoning for why a certain approach was the best.

Our goal is to emulate parts of the nursing process (as described in the previous chapter). However, the assessment and diagnosis phases were too short due to the perfect information. To put more emphasis on these phases, we decided to implement more decisions into the game which required reasoning about contextual factors.

One of these additions was to obscure some factors by requiring the player to measure them first. The player could identify whether a factor was present, but had to use measurement tools separately from the timeline to determine the state of the factor (see Figure 5.5 for an example). This also allowed us to add more complexity by having certain measurements be unnecessary for the solution of a level, providing us with a way to determine whether a player was actually reasoning about contextual factors or not (as we could see whether the player measured useless factors). All measurable factors were chosen to feel at least somewhat logical to measure, i.e. illness could mean a fever (temperature) and the fact that trying to move a sleeping patient is not a very good idea.

Initially, a player could measure as much as they wanted per level. A small pilot experiment showed that this turned out to nullify the reason for adding the measurements in the first place as players simply measured everything all the time. As they were not punished for doing so, this kind of behaviour was implicitly stimulated.

This presented an interesting dilemma: either you limit the amount of measurements and punish players severely when measuring the wrong factor as they are not allowed to measure anything else; or you allow them to measure everything they want and lose some of the educational gains.

In the end we settled for the first option with a slight twist: after making an error due to a wrong timeline, the player was allowed to measure again. In theory, this would still allow players to measure everything. In practice however, this turned out to be less of a problem than expected.

Contextual factors

The prototype only had one type of contextual factor: the weight of the egg (small, medium and, large). The weight was only given in the introductory part of the level, after which it could not be accessed again. To make sure that the player could identify the weight in the level itself if necessary, the weight was made visible through the size of the egg.

To make the game more complex and allow for more possible levels, additional factors were designed, including possible interactions between factors. The factors that were added are:

1: Terrain

All tools were equipped with wheels, meaning terrain could play a role in whether a particular type of tool could be used or not. The three types (grass, gravel, and mud) each have a list with tools which are still usable, with mud being the only type of terrain which prohibited all tools (and thus requiring manual repositioning), gravel only being traversable by foot and with a wheelchair, and grass being traversable by anything.

This factor was chosen to help players eliminate which movement tool had to be used. It allowed us to create an extra decision as well if the terrain was passable for multiple tools.

2: Illness and Temperature

Temperature was added as one of the aforementioned measureable factors. To indicate that temperature might be relevant, an egg would be bright green instead of the bluer colour it

normally has. The player can use a thermometer to determine the temperature, which could be either 30 degrees Celsius or 70 degrees Celsius. The latter case would require the player to put on heat-resistant gloves to resist the skin-scorching heat.

3: Asleep and Emotional State

Another measurable factor was emotional state. This was made visible by having three large 'Z's above the egg, indicating the egg is sleeping. While it may seem odd for an egg to have an emotional state, this was done more so to give the egg a more humane touch as well as provide an intuitive measurement. All measurable factors were chosen to feel at least somewhat logical to measure, i.e. illness could mean a fever (temperature) and trying to move a sleeping patient is not a very good idea.

The two possible emotional states could be either anger ("you had the nerve to wake me up?!") or fear ("Where am I? What's happening?"). Anger requires the player to put on working gloves as a safety measure, fear rules out the possibility of using a hoist. The fact that anger and temperature have an overlapping consequence was intentional: it gives the player the possibility to reason that they do not need to know the temperature of the egg if they are going to wear gloves regardless of the outcome of that measurement.

4: Brittleness

This factor was added as 'test' factor, only to be introduced before the final three levels. This was done to introduce more complexity and to see whether players would be able to reason in terms of contextual factors when they would interact with each other.

An egg can be either brittle or sturdy, each only having meaning in combination with other measureable factors. An angry and brittle egg prohibits the use of a hoist due to safety concerns. A brittle egg cannot be transported over gravel using a wheelchair (again due to safety concerns). A scared but sturdy egg exempts the rule of no hoist allowed.

Feedback

Initially, the game had no feedback other than letting the player know whether a selection of actions was correct or not, having the first wrong action be highlighted red in case of an error. As described in the previous chapter and the instructional design, this is not enough. We implemented a simple pop-up screen that appears after a wrong action was found.

The message of the feedback could relate to two things, depending on which version the player played; either the correct answer was given (corrective feedback) or the player was hinted at the contextual factor responsible for the answer being wrong. These two versions of feedback were designed in order to answer one of our research questions.

At the end of each level, the player received more general feedback on his or her performance depending on their actions during the completed level.

Levels

As stated earlier, the prototype had only one level. This was expanded upon using the instructional blueprint gained through applying the 4C-ID method. The final three levels of the game included the measurable factor brittleness in order for us to test the player.

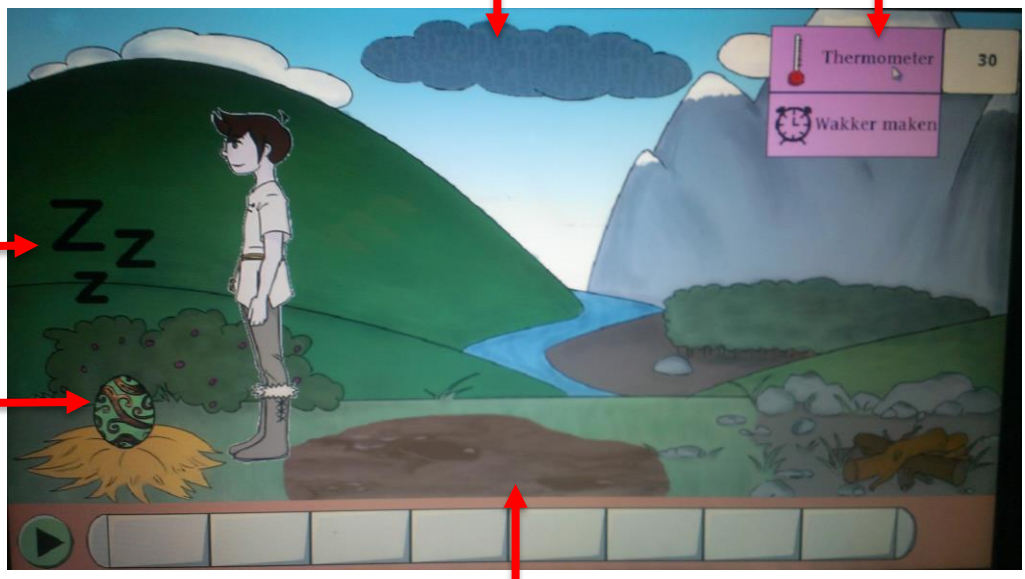
Level design for the second task class came down to exhausting most combinations of factors that did not have an immediate obvious answer. To help the logical reasoning and increase the possible interesting combinations, we added some weather indicators which could also indicate a factor which otherwise would have to be measured. A sunny sky would indicate that the egg is 70 degrees Celsius and dark clouds indicate that the egg is fearful (see Appendix B for a full overview).

Conclusion

With the CTA pinpointing the right problem definition and 4C-ID providing a good instructional design structure, the prototype with which we started is now ready for the experiment. Figure 5.5 provides an overview of some of the changes using one of the levels as example.

The weather can indicate what the value is of one of the obscured contextual factors. The cloud tells the player that the egg is scared, leaving only the egg's temperature as a missing value.

The implemented measurement tools. In most levels the player has only these two options and can choose only one.



The colour of the egg is bright green, indicating the egg is sick. The Z's indicate that the egg is sleeping. Both of these indicators are there to represent obscured contextual factors.

The mud on the ground is an example of 'terrain': one of the added contextual factors. It determines what kind of tools the player can use.

Figure 5.5: An overview of level six. The player has correctly analysed that, although both the emotional state as well as the temperature of the egg could determine whether heat-resistant gloves are necessary, the dark cloud already tells him or her that the egg will be scared. This would mean that the egg cannot be moved using a hoist. However, that could not have been the solution in the first place, as the terrain is muddy: preventing the use of any tools with wheels. Thus the player knows that the temperature has to be taken to see if the egg is too hot to touch without heat-resistant gloves. The egg is 30 degrees Celsius, indicating that the player does not need the gloves and can carry the egg to the other side.

6. Method: Experiment

This section describes the revised research questions, the control and test groups, the participants that took part in the experiment, the manual used to explain the game, the procedure of the experiment, and the measurements taken for later analysis.

Sharpening research questions

Now that the game is redesigned, we can sharpen the research questions to also describe *how* we will measure an improvement in performance. Each change is briefly discussed below:

1. *Do participants who receive reflective formative feedback on their errors perform better at determining the correct contextual factor to measure than participants who receive corrective feedback?*

We want to know whether the participants who receive the reflective feedback become better at analysing new situations. We do not look into the immediate effect of the feedback on whether the test group's subsequent attempts are closer to the solution than the control group, as it is difficult to tell whether a better or worse action sequence in the second attempt is purely due to understanding the feedback. Subsequent measurements provide a clearer picture.

2. *Do participants with experience in playing games perform better at determining the correct contextual factor to measure than participants who have no experience?*

Same as above, but in this case we want to see whether having prior game play experience makes it easier to learn how to reason with the contextual factors, as the game play aspect is likely to be picked up more easily. We also want to see whether having prior game play experience influences the amount of attempts needed to solve the levels as well as the time each level, attempt, and measurement took. We expect that participants with prior game experience will be faster at the game as they will probably adapt more easily to, and interact with, the interface.

3. *Do participants become better at measuring the correct contextual factor throughout playing the game, and if so, does the test group learn better than the control group?*

Lastly, we want to know whether the specific design process used for this thesis leads to an effective learning experience. In answering this question we again mainly focus on whether participants become better at measuring the correct contextual factor at the first attempt of a level, as this is the clearest representation of whether they become better at analysing the situation.

Measurements

Time measurements were taken of each attempt and level, as well as the time participants took to read feedback. These were taken to be able to answer the question whether the participants with game experience were faster at the game than non-experienced participants (this could be an indication that participants with experience in gaming are better at the game: Mayer et al., 2014).

The first measurement of each attempt at each level was measured to see whether participants became better at reasoning with the contextual factors in subsequent new situations if the feedback pointed them towards it instead of providing a direct answer.

The amount of attempts per level was also taken to see whether participants of the test group need less attempts due to the feedback pointing them towards reasoning about the problem beforehand. It is likely that this will make them think longer on each attempt while in turn they would need less attempts.

Participants

The participants included mostly students recruited on the faculty grounds with 33 participants in total (10 women, 23 men). Participants' age ranged between 19 and 61, *M* = 29,76 years, *SD* = 11.34. Participants were randomly assigned to either the test group (reflective feedback) or the control group (corrective feedback) resulting in 17 participants assigned to the test group and 16 to the control group.

Materials

Location

All experiments were held inside the faculty's experiment room which provides separate booths with each a macbook, monitor, keyboard, and mouse.

Survey

After filling in the form of consent, participants were assigned to a computer and received a survey. The survey consisted of the following items:

1. Age;
2. Gender;
3. Educational degree;
4. Previous work experience and education in healthcare;
5. Time spent playing games per week;
6. Preferred game genres.

Aside from general items such as age, previous work and educational experiences in the healthcare sector were surveyed to be able to analyse prior domain-knowledge effects.

The last two questions were added to analyse prior gaming-experience and effects of game genres. Question five was implemented as a five-point scale, ranging from 'zero hours per week' as the lower end of the scale to 'more than 10 hours per week' as the higher end (also used by Clark et al., 2011). This ties into the research question about prior gaming experience and its effect on measurements and play time (i.e. people with game experience are better at playing the game than non-gaming participants). Finally, the item on preferred game genres is there to see whether having a lot of experience in the same genre as the designed game (puzzle genre) has a positive effect on measurements and play time.

Manual

The instruction manual was there to explain the goal of the game, its rules (e.g. the contextual factors), interface, and the possible ways in which the egg could be moved (i.e. manually, using a wheelchair, or using a hoist). The options would only be described plainly, the specific action sequences necessary to perform one of the three options were not given. The reason for this goes back to the background literature, where we discussed the nursing cycle of Burns, O'Donnell, and Artman (2010). Although the full cycle is used here, the focus of the manual was mostly on the rules for analysis and diagnosis.

All parts were explained using text accompanied by images, for example showing the contextual factors and describing their implications (see Appendix B).

Procedure

Upon entering the room, participants would be greeted and asked to read and fill out a form of consent. Then, the goal of the game and the core mechanic (i.e. programming a set of moves) was explained, after which the player was appointed to one of the computers and asked to fill in

the survey before starting the game. The participant was notified of the fact that, were any question to arise, the participant was free to notify and ask the researcher.

Participants could only use the manual during the first three levels to familiarize themselves with the rules, controls, and interface. After the third level, the participant would notify the experimenter to pick up the manual. The restriction of only having the manual the first three levels was done to prevent noise in the time measurements, as it was impractical to determine how often a participant was using the manual and log that time as well.

After finishing all twelve levels, the participant was thanked and given a small treat (candy). Any questions regarding the experiment were answered by the researcher.

7. Results

This section describes the results of the experiment. The analysis will be focused on whether there's a difference between the control and test group's ability to measure the right contextual factor (*Correct_Measurement*). This variable can be either 'true' or 'false', as it indicates whether the participant's first measurement per level was used on the correct contextual factor. Another possible indicator is *Attempts*, the variable which tells us how many tries a participant needed to find the correct solution of a level.

Other analyses will focus on the influence of game experience on *Correct_Measurement* and the following time related variables:

- *Average_Time_Attempt*: the average time (in seconds) taken per attempt per level per participant;
- *Average_Time_Feedback*: the average time (in seconds) taken to read feedback per level per participant;
- *Time_Level*: the total time (in seconds) taken per level per participant;
- *Measurement_Time*: the time (in seconds) taken to choose which factor to measure during the first attempt per level.

Because the first three levels were there to help the participant get to know the game, they were omitted from the analyses. Adding these three levels would reintroduce the noise we wanted to keep out of our data as described in the previous chapter.

Research questions

The research questions are repeated here to provide clear goals for the statistical analysis. For each question, the statistical methods that have been used to find an answer for them are mentioned, including the reasoning behind their inclusion.

Research Question 1

Do participants who receive reflective formative feedback on their errors perform better at determining the right contextual factor to measure than participants who receive corrective feedback?

To be able to answer this question a chi-square test is taken to see whether, overall, the test group makes more correct measurements than the control group. The differences in attempts needed per level is also analysed using an ANOVA.

Research question 2

Do participants with experience in playing games perform better at determining the right contextual factor to measure than participants who have no experience?

This question is answered by performing another chi-square test to see whether there are differences between the different categories of *Gaming_Experience*. ANOVA's are performed to see whether there are significant differences between the different experience groups on the variables *Average_Time_Attempt*, *Average_Time_Feedback*, *Time_Level*, *Measurement_Time* and *Attempts*.

Research question 3

Do participants become better at measuring the correct contextual factor throughout playing the game, and if so, does the test group learn better than the control group?

To answer this question, a repeated measures ANOVA was used to see whether there were significant within-subject differences throughout the levels. Version was passed on in this test as a between-subject variable to see if there were any differences between the test group and control group. We account for the effect of Gaming Experience by passing it on as a covariate to make sure we get pure statistical information on the effectiveness of our game.

Descriptive statistics

Amount of correct measurements

Figure 7.1 shows a drop-line graph of the ratio of correct measurements per level of both the test and the control group. Aside from levels four, six, and eight, the differences are small and point towards no significant difference in performance. Also note the relatively low amount of correct measurements in levels eight and eleven. Overall, the trend seen in this graph tells us that, while the control group does seem to learn somewhat, there is little to no learning effect. Our tests will have to see whether this is true.

Non-categorical data

Table 7.1 shows some descriptive analysis such as the mean and standard deviation for variables we are interested in for our second research question. The data is sorted on only the different versions, but one can see that both have a large standard deviation for some of the variables related to time. This could indicate that these differences could be explained by other differences such as gaming experience.

Overall difference in performance

R.Q. 1: Difference between feedback groups

Because level seven did not require any measurement to be done, this level will also be excluded from the analysis. To see whether being in one of the two feedback versions leads to a participant making more correct measurements, a Pearson chi-square test was performed. The results show that both groups are independent from one another $\chi^2(1, N = 297) = 1{,}330$, $p = .249$. This means that the groups did not differ significantly in the number of correct measurements.

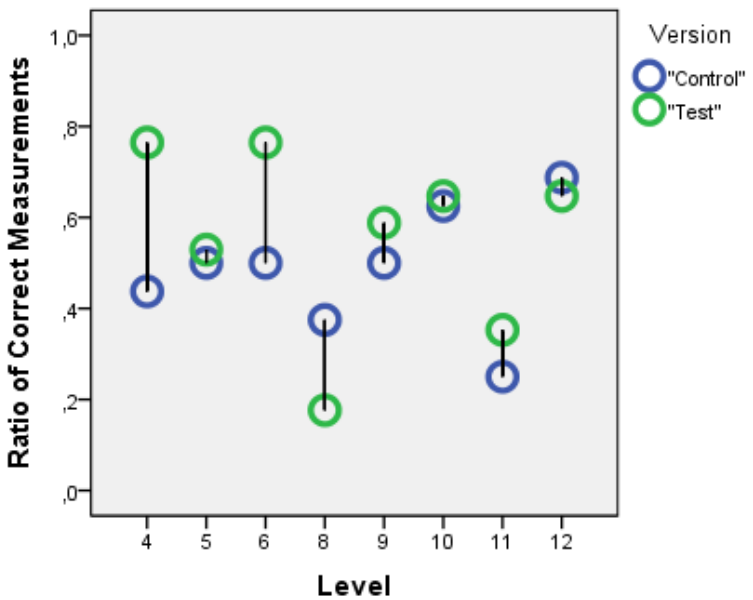


Figure 7.1: A graph showing the ratio of correct measurements per level for the test and control group. Level seven is omitted as there was no measurement in that level. Aside from levels four, six, and eight, the differences are very small.

Table 7.1: Descriptive statistics of all the non-categorical variables. Note the large standard deviations for average time taken per attempt (Av. Time per Attempt), time taken per level (Av. Time taken per Level), and the time taken per measurement (Av. Time taken to Measure).

Variable	Version	N	Mean	Std. Deviation
Attempts	Test	192	2,38	1,62
	Control	204	2,57	2,51
Av. Time per Attempt	Test	192	44,42	26,45
	Control	204	40,67	22,10
Av. Time reading Feedback	Test	123	5,07	7,35
	Control	107	4,70	5,67
Av. Time taken per Level	Test	192	134,14	73,86
	Control	204	136,17	107,90
Av. Time taken to Measure	Test	125	16,12	18,67
	Control	146	14,00	15,40

Comparing the means of each group's attempts per level required a Brown-Forsythe robust test of equality of means due to a violation of the homogeneity of variances. The test reveals no differences in the average amount of attempts used per group per level (p = .796).

R.Q. 2: Difference between categories of Gaming_Experience

Looking at the amount of correct measurements between experience groups, a Pearson chi-square test reveals that the groups are not independent from one another $\chi^2(1, N = 297) = 4{,}948$, $p = 0{,}026$.

Table 7.2 shows the results of a Brown-Forsythe test that was performed to see whether there were differences between the categories of *Gaming_Experience* for the variables *Average_Time_Attempt*, *Average_Time_Feedback*, *Time_Level*, *Measurement_Time* and *Attempts.* Significant difference were found between the categories for the average time taken per attempt (F = 7,13, p < 0,000), average time taken per level (F = 3,24, p = 0,013), and the average time taken to measure one of the two obscured contextual factors (F = 2,95, p = 0,022).

Difference in learning

Overall

Looking only at the *Correct_Measurement* variable, a repeated measures ANOVA was performed to see whether participants became better at measuring the correct contextual factor on their first try.

Mauchly's test indicated that the sphericity assumption had been violated (p = 0,006), so a Greenhouse-Geisser corrected test was used (ε = 0,68). No significant within-subject effect was found throughout the levels, F(4,77, 143,01) = 1,229, p = 0,288.

Table 7.2: Results of the Brown-Forsythe test showing that there are significant differences between the categories for the variables Attempts, Av. Time per Attempt, and Av. Time reading Feedback.

Variable	Asymptotical F-value	Degrees of Freedom	p-value
Attempts	2,03	4	0,093
Av. Time per Attempt	7,13	4	**0,000**
Av. Time reading Feedback	0,41	4	0,799
Av. Time taken per Level	3,24	4	**0,013**
Av. Time taken to Measure	2,95	4	**0,022**

Table 7.3: Standardized residuals of the Pearson chi-square test of Correct_Measurement and the categories of Gaming_Experience.

			Gaming Experience Categories (hours per week)					Total
			0	**1 to 3**	**4 to 6**	**7 to 10**	**10+**	N
Correct Measure-ment	No	N	16	34	16	33	27	126
		Std. Res.	**-2,1**	-0,1	0,2	1,2	0,9	
	Yes	N	40	38	16	23	21	138
		Std. Res.	**2,0**	0,1	-0,2	-1,2	-0,8	
Total		N	56	72	32	56	48	264

R.Q. 3: Test and control group

The grouping variable *Version* showed no significant result (p = .588).

Post-Hoc Analyses

Categories of Game_Experience'

The significant result of the Pearson chi-square test was followed up by looking at the standardized residuals seen in table 7.3. The residuals show that the group with no gaming experience caused the significant effect of the Pearson chi-square test.

The significant results found in the ANOVA's in table 7.2 are looked into more deeply by calculating their correlations, as seen in table 7.4. *Gaming_Experience* negatively correlates with both *Average_Time_Attempt* and *Time_Level*.

The line plots of Figure 7.2 clarify the correlations more thoroughly, showing that participants with more gaming experience generally play faster.

Table 7.4: The correlations between the average time taken per attempt, average time taken per level, average time taken to measure, and the categories of gaming experience.

		Av. Time per Attempt	**Av. Time per Level**	**Av. Time taken to Measure**
Gaming Experience	Pearson Correlation	**-0,202**	**-0,135**	-0,115
	p-value (2-tailed)	0,000	0,020	0,097
	N	297	297	211

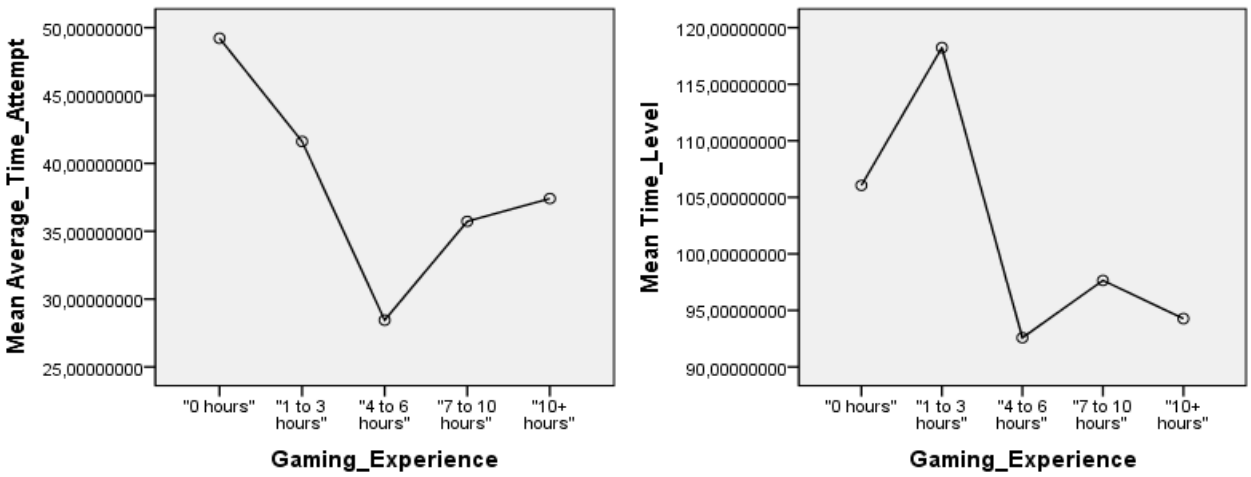


Figure 7.2: The left line plot shows the average time taken per attempt of each category of gaming experience. The right line plot shows that these lower averages per attempt lead to large differences in average time taken to finish a level. Especially between the participants with no game experience and the participants who play games for more than four hours per week.

8. Discussion

This section discusses the methods used for this thesis, which include the results of the experiment, the CTA used for preliminary research, the 4C-ID method used to (re)design the game, the game itself, and the experiment setup itself.

Interpreting results

Test group and control group

The test group did not perform significantly better than the control group nor was there a significant learning effect from the game alone. The latter indicates that the differences between the performance of the participants over the levels can mostly be explained by a participant's gaming experience and other personal factors.

A possible reason for the lack of a learning effect could be the fact that the most prominent part of the game play was creating an action sequence, whereas the focus of our learning goal was more geared towards the analysis beforehand. Thus there was a mismatch in the amount of importance the analysis had for our learning goals and our in-game goals, also making the reflective feedback less relevant. This is something we will discuss more thoroughly in the 4C-ID section.

Sadly, there being no difference between groups suggests that the feedback type did not impact the learning nor the performance of the participants. This answers both our first and our third research question with a resound 'no'. Formative feedback did not train participants to become better at measuring contextual factors than summative feedback. This is contrary to the literature we could find on differences in feedback (e.g. Shute, 2008; Adams & Clark, 2014), indicating that our integration probably did not provide the intended reflective moment.

We believe that part of this is due to the disconnection between *when* a measuring error is made and *when* the participant receives feedback about it. The feedback in our game was only presented after the participant created a sequence of actions and pressed the 'go' button. In the test group version, the feedback is more tailored towards an error the participant made much earlier on (depending on the time it took to make the first sequence). This time gap could explain the lack of significant effectiveness. Literature suggests that our the reflective feedback would have worked better if it had been closer to the actual error (Shute, 2008).

Small indications of feedback effect

We mention a lack of *significant* effectiveness, because there are several non-significant effects which do point toward a *possible* effect. Figure 8.1 shows a bar graph of the types of measurements made:

- the participant made the correct measurement on the first attempt;
- the participant made the correct measurement on one of the later attempts (i.e. the user noticed their mistake and wanted the right information);
- The participant made an incorrect measurement and did not measure again; and
- The participant made no measurement at all.

One can see that the control group did seem to perform worse (albeit not significantly), but what is even more interesting is the fact that there is a substantial difference between the two groups on the second category of measurements. This could indicate that the feedback did stimulate participants to go back and rethink *why* they measured the wrong factor.

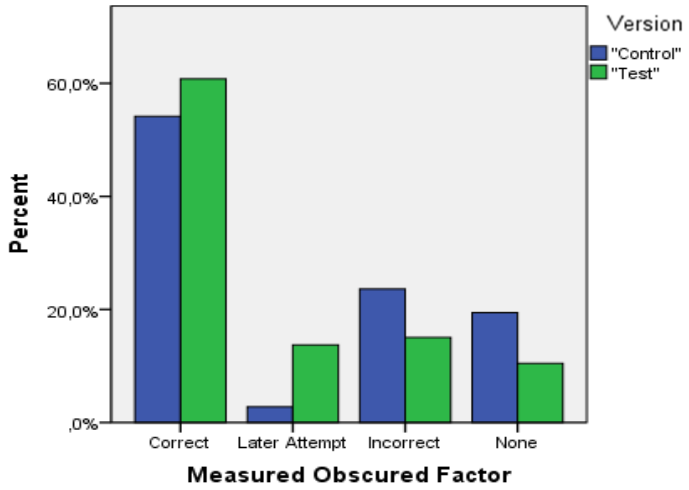


Figure 8.1: A clustered bar graph depicting the percentages for each of the four possible measuring categories. The second category shows a substantial difference, indicating that the reflective feedback possibly caused participants to reanalyse the situation.

However, it could also mean that the participants mindlessly clicked the other measurement as they were reminded to look at the contextual factors.

Another possible indication of an effect of the feedback can be found by looking at the time spent on the first measurement per level (see Figure 8.2). Here we can see that in most levels the test group takes more time before they measure. This could indicate that they reason more about their choice, but the effect is not statistically significant.

These effects would probably have been more prominent, and possibly significant, if there had been more time to train and if, as mentioned earlier, the analysis part of the game play had been more important towards achieving the in-game goals.

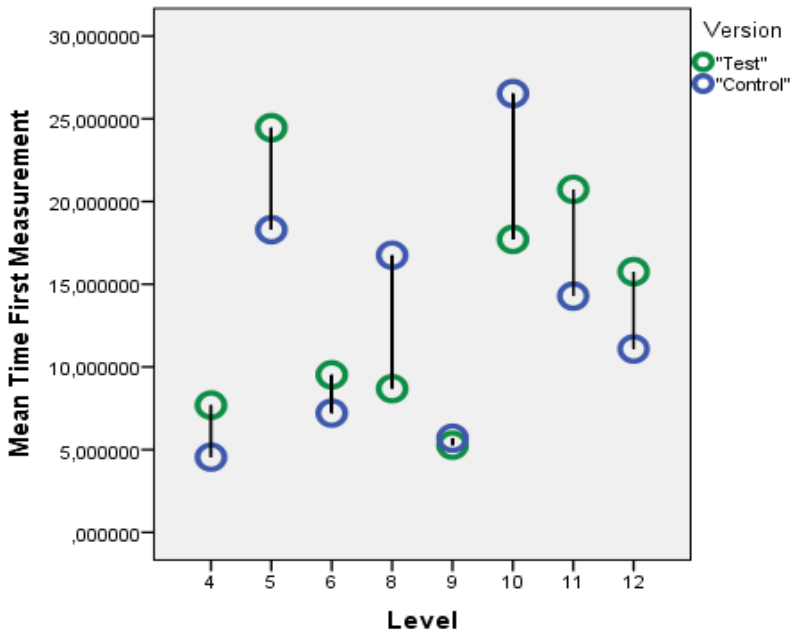


Figure 8.2: Drop-line graphs showing the difference in average time spent on the first measurement between the test and control groups. Level seven is again omitted due to the fact that there is no measurement necessary in that level. In most levels the test group takes more time for their first measurement, possibly indicating more thinking time..

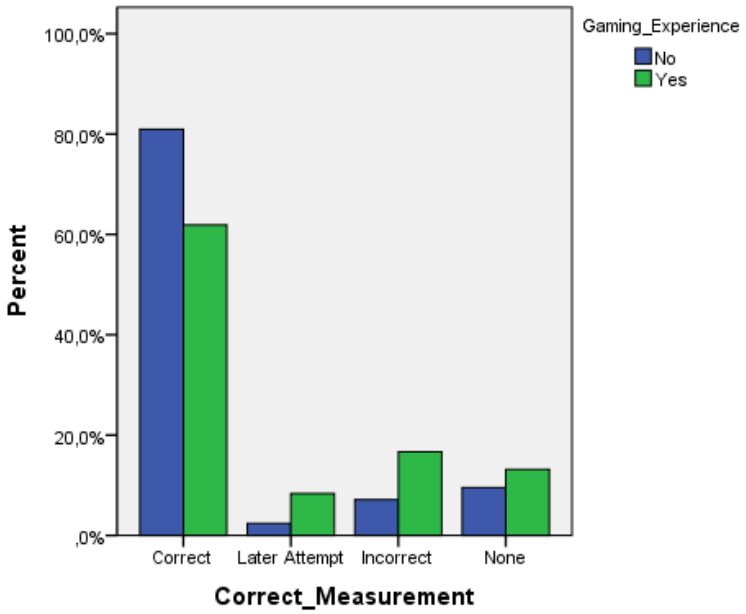


Figure 8.3: All Gaming_Experience categories with more than zero hours of playing games per week are collapsed into one category to show that participants with gaming experience make more measuring mistakes than participants without experience.

Categories of Gaming_Experience

The Pearson chi-square test revealed that there was a dependence between at least one of the pairs of categories. The standardized residuals then revealed an interesting difference: the group with no gaming experience was, overall, *better* at measuring the correct contextual factor at the first try. This answers are second research question with a 'no'.

There are multiple interpretations possible here:

1. People with gaming experience are challenging themselves to play the game with as little measurements as possible, causing them to not measure at all. This does not seem to be the case, as Figure 8.3 shows that participants with gaming experience just performed worse *overall*. The fourth category (no measurement made) also shows no huge difference between the two groups.
2. People with gaming experience play the game faster than their non-experienced counterparts, which causes them to make more measuring errors.
3. The reflective feedback confuses participants with gaming experience more than participants with no gaming experience.

The second interpretation is based on the correlations found in the post-hoc analyses, which showed negative correlations between *Gaming_Experience* and both *Average_Time_Attempt* and *Time_Level*. This indicates that participants with gaming experience took less time for their attempts and thus less time per level. However, *Measurement_Time* was not found to be a significant negative correlation (only -.115, p = .097), thus we cannot conclude that they took less time for their measurement and that this would cause them to make more errors.

The last interpretation is based on the idea that the reflective feedback is not as relevant to the in-game goals as we had wanted. The significant worse performance could be due to the fact there are significantly more participants with gaming experience in the test group than in the control group (ANOVA, F=23,70, p < 0,001). Fifteen of the seventeen participants that played the reflective feedback version had gaming experience (as opposed to eleven of the sixteen in the test group), which could indicate that the participants with gaming experience made little use out of the reflective feedback.

This is strengthened by the explanation that, as both groups started with barely any knowledge of ergonomic safety, the reflective feedback was too complex for the test group. This could also explain why the participants with no game experience performed better, as most of them were in the control group, which received less complex feedback. Literature tells us that experts or more knowledgeable learners have different feedback needs than novice learners (e.g. Shute, 2008), so our results could indicate that we failed to take this into account (i.e. our tutorial was focused too much on teaching game play only).

Cognitive Task Analysis

The CTA method helped a great deal in clarifying the domain of ergonomics in healthcare: the knowledge that is involved, what decisions have to be made and how. For educational game designers it is important to know what the educational goal of your game is going to be. By making sure that SME's are involved in the exploratory part of your design you guarantee a grounding in expert knowledge. This helps a designer make well-thought choices in their goals and mechanics.

We will discuss our experiences with the CTA methods in the sections below:

When to use CTA

Our domain posed some problems for identifying a *cognitive* skill to analyse using CTA, as initial literature research indicated that the problem was mostly behavioural. By interviewing our SME's we discerned some possible underlying skills which we could analyse further using CTA.

The lesson to learn here is rather straight-forward: only use a Cognitive Task Analysis if you are very certain that what you are going to teach is a cognitive skill.

Scope of the game

Another important consideration has to do with establishing the intended scope of the game. By analysing the problems in more detail through a CTA, we discovered that the problem space is actually quite large and consists of many different aspects. We could focus purely on some decision making moments in a small selection of tasks and create a reasonably sized game around it. After determining the problem space through the CTA, one really needs to choose the most salient goals to include into the (final) game.

Performing a CTA as a designer helps identify what skills need training the most. This in turn allows you to identify the most salient learning goal for the game. Salient meaning that achieving that learning goal will have the learner improve on the most crucial skills.

Design space and CTA

Another interesting aspect of using CTA for preliminary research is the influence it has on potential design space. Design space describes the breadth of options and tools a designer has to create a game; a basic example would be the general game play style (e.g. a first-person shooting game or top down strategy game).

In educational game design this design space is often restricted due to what a game aims to achieve in terms of learning and the fact that it has to adhere to (at least some) real-world rules. We initially reasoned that CTA diminishes this space even further, as it provides a more detailed, and thus narrowed down, understanding of what the user needs and thus what the game needs. However, we found that having the data of the CTA at ones disposal increases the *use* one can get out of this limited design space (i.e. it is easier for the designer to understand the consequences of the design choices with relation to the overall goal). A better understanding of the domain allowed us to be more creative within the constraints of the problem.

The lesson we learned here is that more knowledge means more options. The real-world rules are not as much a restriction to the creative process as they are a guiding line to what is possible within the domain.

Benefit of CTA

To round up this section, we believe CTA to be a very important and strong tool for future endeavours in designing educational games which train complex cognitive skills. The array of tools combined with the fact that this method is widely validated makes it a strong choice even when no other non-game related methods are used.

Four Components Instructional Design

After understanding the details of a complex skill, the next step would be to design the instructions to teach that skill. The 4C-ID method was a thorough way to build an educational experience. So thorough that one of the four seemed unnecessary (part-task practice). The sections below describe the problems we encountered whilst using the 4C-ID method for our game's design.

Task classes and their depth

Early on in our use of the 4C-ID model, we used all the information the CTA provided to create a wealth of task classes. These task classes would cover all the possible exceptions and combinations of contextual cues. This quickly turned out to be a bad approach, as an enormous amount of task classes would have had to be made to cover everything. This would also go against our idea of teaching a flexible strategy, as argued in Chapter 3. Instead, we changed our approach by looking at the commonalities between tasks and creating task classes which dealt with teaching about these commonalities, reinforcing recognition in the user.

The guideline here is to make clear, smaller, learning goals per task class and to make sure that each task class' main focus is to achieve that learning goal.

Translating 4C-ID components to game design

Although the 4C-ID model neatly fits the CTA data in the creation of an instructional design, we experienced some difficulty with translating all its components to game design elements. This problem was caused mostly by the fact that we started with a prototype, which pre-emptively determined some of the implementations of the instructional design (these issues will be discussed in its own section more thoroughly).

There are two sides to the translation problem: 1) following an instructional design method without considering why each part is relevant can lead to problematic situations in which something is implemented 'because the method says so'; and 2) Not including all aspects of the method can lead to unexpected results which can be difficult to retrace to the omission of an aspect.

We excluded one part (on-task practice) of the model and thus had to deal with the consequences of the second problem. Looking at the results and the fact that the analysis aspect should have had more game play relevance, we could have used a small sub-game as a way to quickly train the analysis some more at the end of the first task class.

The thing to be learnt here is that, while the 4C-ID model provides a nice structure, one has to be aware that it increases the game's scope as it possibly works best when all components are used.

Continued discussion results

We believe that one of the reasons we failed to see that the analysis aspect of our levels lacked relevance towards the game goals is because of the fact that the 4C-ID model does not deal with the in-game goals. It is the task of the designer to maintain this, making it easier to forget when you purely use the 4C-ID method.

To remedy this, we think that an interesting step to take would be to use existing educational game design frameworks to cover the game play aspect. Arnab et al.'s (2014) Learning mechanics - Game mechanics model for example could be used to denote the distance between the feedback and the error as well as more easily identify the way in which the learning goal relates to the in-game goal. A different possibility would be the activity-based model of Carvalho et al. (2015) which allows levels to be described in both their game play as well as their instructional aspect. The latter describes the game play at the same level as the 4C-ID method, making it an interesting and compatible combination.

It was only after the experiments, while writing this thesis, that we also found a paper which already looked into using 4C-ID for educational game design (Huang and Johnson, 2009). Research that flowed out of this work showed two previous research endeavours looking into 4C-ID and CTA (Rosmalen et al., 2012; Boyle et al., 2012; Lukosch et al., 2012). However, neither of these games used the format presented by Huang and Johnson. Future endeavours should first look into the applicability of Huang and Johnson's work.

Benefits of the 4C-ID method

To conclude this section we describe some of the strong points we experienced while using the model:

1. The general lesson structure the method provides really helps to get a grip on the underlying learning process you want from your educational game. It ties neatly into the process of designing levels, which most games have to go through.
2. It provides a guideline for applying scaffolding, allowing one to reason about the difficulty of levels in terms of the to-be-learnt skill. Existing game design frameworks do not yet provide such formal methods.

Game 'How to help your dragon'

Using a prototype

In hindsight, using an existing prototype has proven itself to be a very limiting factor. However, researching for, designing, implementing, and testing a game about training a complex decision-making task would have been way beyond the scope of this thesis, thus we had to accept this limitation.

We were glad to have at least found a game which had core game play that matched the requirements that came from our preliminary research (although we believe the action-sequence creation was too prominent). As described in our problem instantiation in chapter 3, we wanted to address the nursing problem-solving cycle. The prototype allowed us to do so, after implementing a more detailed diagnosis mechanic (i.e. the measuring of unknown contextual factors).

The core mechanic

While the 'action-programming' mechanic is a good tool to foster thinking before doing, as seen in the example of SURGE: Fuzzy Chronicles in chapter 2, it was too unintuitive at times. Small usability errors, such as only accepting one subsequence of actions even though intuitively the order could be different, led to attention being taken away from the diagnosis aspect of the levels.

It also took up a lot of playtime, so it prevented us from adding more levels in an effort to keep the experiment in a manageable time-range.

An example of this was seen with levels that required both heat resistant gloves to be worn as well as a wheelchair to be used. In the game, the gloves always had to be put on first, but participants did not feel that this order was very necessary. This lead to some groans and overall confusion.

Too little variety

The game's 2D look at a landscape and the fact that one could only go to the left and the right also hampered the variability in scenario's. The option to add challenges such as dealing with limited space, patient mobility, and other more close-to-home examples was limited due to the simplicity of the prototype combined with our lack in art development skills.

The length

Due to the lack of possible variety, the possible amount of diverse scenarios was limited as well. This made it difficult to design more levels while maintaining the requirements of the 4C-ID method (e.g. strong variety, gradual decrease in scaffolding, and increasing complexity). This could have also influenced the learning effectiveness of the game.

Feedback system

There were also some issues concerning the way the formative feedback system operated. For example in level eight, the feedback system provided the wrong feedback when participants made an error in the ordering of the heat resistant gloves and wheelchair. This would have been caught if more people had been used during the pilot experiment.

Lessons learnt

A new approach to teaching the flexible skillset should involve more focus on the diagnosis aspect of the problem-solving skill. This way the scenarios can be more diverse, which would possibly allow for part-task training to be reintroduced again, as there are more overlapping subskills. In the current version of the game, the scenario pool was so shallow that part-task training would have been either extremely dull or it would have taken up space regular levels would have used.

Intuitive, user-centred design is also needed. The lack of playtesting became clear when the small, unintuitive (and unknown to the designers) design choices caused confusion and frustration amongst the participants. More user feedback in terms of image sorting and describing would greatly help a subsequent version avoid the now known pitfalls.

Experiment

Overall, the experiments were administered without any major problems. There are, however, some issues more related to the method that we do wish to discuss.

Testing for an effect

The setup of the experiment made it hard for us to find a learning effect, as there were only a limited number of levels and thus not that many data points to use. Ideally, we would have used a pre and post test to determine the learning effect of the game and knowing more about the base level of each participant. However, we were unable to find a validated test for our purposes (i.e. recognizing and using contextual factors for diagnosing in a nursing context) and the scope of the project prohibited us from developing our own.

Erroneous choices in level design

In applying the 4C-ID method, we stretched the level diversity too far by allowing some levels to have multiple correct ‘implementation’ solutions but only one measurement (‘diagnosis’) solution. For example, level nine could be solved using any of the three possible movement strategies (no tools, wheelchair, and hoist), but if you wanted to know the most fitting strategy one would still have to measure the correct contextual factor. While these kinds of setups could lead to interesting data (do people who measure better know which approach to use here?), it only made the measurement too trivial and thus the level too easy.

Another example is level seven in which participants did not have to measure anything. While providing some nice difference in game play, it mostly meant that we had one less data point per participant to see whether they learnt to measure correctly.

These kinds of problems could be avoided if the experiment design and its most relevant variables are known during the game’s design phase.

9. Conclusion

In this study we investigated the possibility of using non-game related methods to create an educational experience in a video game. To test this, we analysed a complex problem solving skill in the domain of nursing and adapted an existing prototype of an educational game to train this skill. The educational game was also be designed with the cognition of the user in mind, making sure that the player was trained both in game play and in domain knowledge.

The first non-game related method used, the cognitive task analysis, proved to be a valuable tool for gaining proper understanding of the domain and the problem to be solved. This method is strongly recommended to educational game designers in need of a tool for finding knowledge on teaching a cognitive, complex, skill.

The second non-game related method, four component instructional design, was more difficult to apply fully. However, it provided a structural method to introduce the educational aspect in level design. This allowed us to systematically scaffold the challenges provided by the game play as well as the domain knowledge. The problems we ran into mostly rose from the fact that the 4C-ID method only creates the instructional design of a training (i.e. educational game), not the way in which it is trained. This made us lose track of the role of the game play in achieving the learning goal, leading to some game mechanics not being as useful for the learning goal.

The game's two versions were there to see if we could increase the learning effectiveness of the game if the feedback was there to promote reflection on the reasoning ('diagnosis') aspect of the complex skill. This difference did not lead to an increase in learning effectiveness, answering our first and second research questions with a 'no'.

We now know that these two non-game related methods have interesting possibilities for the field of educational game design. Thus it is in our best interests that we keep in mind that other design methods exist as well, and that the field of educational game design is not relegated to game design-specific frameworks only.

Future research could look into the following questions:

1. Does designing an educational game from scratch using the 4C-ID method lead to an effective learning experience (i.e. instead of applying it to a prototype)?
2. When designing an educational game from scratch, which methods of a cognitive task analysis help in determining the kind of core mechanic that is needed for the game to be effective?
3. How can existing educational or instructional game design frameworks be used to reinforce the 4C-ID method, making sure the learning goals and in-game goals are more closely aligned?
4. How can we compare educational games that are developed using the 4C-ID method and does this provide insight into more general principles of effective educational game design?
5. How can we properly compare educational game design processes to discern which ones are effective?

To conclude, the fact that existing, non-game related, methods are not easily usable for educational game design tells us that games have a side to them that is inherently different from other educational media. As such, it is our job to determine how to use these differences to make a new tool for our toolbox of educational media.

10. References

1. Adams, D. M., & Clark, D. B. (2014). Integrating self-explanation functionality into a complex game environment: Keeping gaming in motion. Computers & Education, 73, 149-159.
2. Ainsworth, S. (2013). Screenshot of Zombie Division. Retrieved from http://www.nesta.org.uk/blog/are-digital-games-valuable-tool-learning
3. Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., Freitas, S., Louchart, S., Suttie, N., Berta, B., & De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, *46*(2), 391-411.
4. Baba, D. M. (1993). Determinants of video game performance. *Advances in psychology*, *102*, 57-74.
5. Baddeley, A. D. (1997). *Human memory: Theory and practice*. Psychology Press.
6. Barab, S. A., Scott, B., Siyahhan, S. Goldstone, R., Ingram-Goble, A., Zuiker, S., & Warren, S. (2009). Transformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education and Technology*.
7. Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological bulletin*, *128*(4), 612.
8. Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a taxonomy linking game attributes to learning an empirical study. *Simulation & Gaming*, *43*(6), 729-760.
9. Bellotti, F., Berta, R., & De Gloria, A. (2010). Designing Effective Serious Games: Opportunities and Challenges for Research. iJET, 5(SI3), 22-35.
10. Boyle, E., Van Rosmalen, P., MacArthur, E., Connolly, T., Hainey, T., Johnston, B., Ger, P., M., Manjón, B., F., Kärki, A., Pennanen, T., Manea, M., & Starr, K., (2012). Cognitive task analysis (CTA) in the continuing/higher education methods using games (CHERMUG) Project. In *Proceedings of the 6th European Conference on Games Based Learning: ECGBL* (p. 63). Academic Conferences Limited.
11. Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.
12. Bronkhorst, B., Ten Arve, A., Spoek, M. & Wieman, D. (2014) Gezond werken in de zorg. Onderzoek naar fysieke en psychosociale arbeidsbelasting onder zorgmedewerkers. Erasmus University Rotterdam, Stichting IZZ, Apeldoorn.
13. Burns, H. K., O'Donnell, J., & Artman, J. (2010). High-fidelity simulation in teaching problem solving to 1st-year nursing students: a novel use of the nursing process. *Clinical Simulation in Nursing*, *6*(3), e87-e95.
14. Carvalho, M. B., Bellotti, F., Berta, R., De Gloria, A., Sedano, C. I., Hauge, J. B., Hu, J., & Rauterberg, M. (2015). An activity theory-based model for serious games analysis and conceptual design. *Computers & Education*, *87*, 166-181.
15. Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, *62*(2), 233-246.
16. Clark, R. E., Feldon, D., van Merriënboer, J. J., Yates, K. and Early, S. (2008) "Cognitive task analysis" *Handbook of research on educational communications and technology*, Vol. 3, pp 577-593.
17. Clark, R. C., & Mayer, R. E. (2011). E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning. John Wiley & Sons.
18. Clark, D. B., Nelson, B. C., Chang, H.-Y., Martinez-Garza, M., Slack, K., & D'Angelo, C. M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers & Education*, *57*(3), 2178–2195.
19. Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2015). Digital games, design, and learning a systematic review and meta-analysis. *Review of educational research*, 0034654315582065.

20. Clark, R. E., Yates, K., Early, S., & Moulton, K. (2010). An analysis of the failure of electronic media and discovery-based learning: Evidence for the performance benefits of guided training methods. In K. H. Silber & R. Foshay (Eds.), *Handbook of training and improving workplace performance: Vol. I. Instructional design and training delivery* (pp. 263–297). Somerset, NJ: Wiley.

21. Cuban, L., & Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Harvard University Press.

22. Dickey, M. D. (2011). Murder on Grimm Isle: The impact of game narrative design in an educational game-based learning environment. *British Journal of Educational Technology*, *42*(3), 456-469.

23. Druckman, D., & Bjork, R. A. (1994). *Learning, remembering, believing: Enhancing human performance*. Washington, DC: National Academy Press.

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

25. Gunter, G. A., Kenny, R. F., & Vick, E. H. (2008). Taking educational games seriously: using the RETAIN model to design endogenous fantasy into standalone educational games. *Educational Technology Research and Development*, *56*(5-6), 511-537.

26. Habgood, M. P. J., Ainsworth, S. E., & Benford, S. (2005). Endogenous fantasy and learning in digital games. *Simulation & Gaming*, *36*(4), 483-498.

27. Habgood, M. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences*, *20*(2), 169-206.

28. Hattie, J., & Yates, G. C. (2014). How knowledge is acquired. In *Visible learning and the Science of How We Learn*, 1st ed. New York: Routledge, ch. 16, pp. 146-156.

29. Holbert, N., & Wilensky, U. (2012). Representational congruence: Connecting video game experiences to the design and use of formal representations. *C. Kynigos, JE Clayson, & Yiannoutsou (Eds.), Constructionism: Theory, Practice and Impact*, 370-739.

30. Huang, W. D., & Johnson, T. (2009). Instructional game design using cognitive load theory. In *Handbook of research on effective electronic gaming in education* (pp. 1143-1165). IGI Global.

31. Jalink, M. (2014). *Validation of a video game made for training laparoscopic skills* [S.l.]: [S.n.]

32. Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological bulletin*, *89*(1), 47.

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

34. Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay?. *Computers & Education*, *51*(4), 1609-1620.

35. Ke, F. (2009). A qualitative meta-analysis of computer games as learning tools. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education* (Vol. 1, pp. 1–32). Hershey, PA: Information Science Reference.

36. Ke, F., & Grabowski, B. (2007). Gameplaying for maths learning: cooperative or not?. *British Journal of Educational Technology*, *38*(2), 249-259.

37. Kickmeier-Rust, M. D., & Albert, D. (2010). Micro-adaptivity: Protecting immersion in didactically adaptive digital educational games. *Journal of Computer Assisted Learning*, *26*(2), 95-105.

38. Kirschner, P. A., & Van Merriënboer, J. J. G. (2008). Ten steps to complex learning: A new approach to instruction and instructional design. In T. L. Good (Ed.), *21st century education: A reference handbook* (pp. 244-253). Thousand Oaks, CA: Sage.

39. Lukosch, H., Van Bussel, R., & Meijer, S. (2012). A game design framework for vocational education. *International Journal of Social and Human Sciences*, *6*(1).

40. Mayer, R. E. (2001). *Multimedia Learning.* Cambridge: Cambridge University Press.

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

42. Mcquiggan, S. W., Rowe, J. P., Lee, S., & Lester, J. C. (2008). Story-based learning: The impact of narrative on learning experiences and outcomes. In *International Conference on Intelligent Tutoring Systems* (pp. 530-539). Springer Berlin Heidelberg.

43. Plass, J. L., O'keefe, P. A., Homer, B. D., Case, J., Hayward, E. O., Stein, M., & Perlin, K. (2013). The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation. *Journal of Educational Psychology*, *105*(4), 1050.

44. Reiser, R. A. (1987). Instructional technology: A history. In R. M. Gagné (Ed.) *Instructional technology: foundations*, 11-48. Routledge.

45. Van Rosmalen, P., Boyle, E. A., Nadolski, R., van der Baaren, J., Fernández-Manjón, B., MacArthur, E., Pennanen, T., Manea, M., & Star, K. (2013, October). Acquiring 21st Century Skills: gaining insight into the design and applicability of a serious game with 4C-ID. In *International Conference on Games and Learning Alliance* (pp. 327-334). Springer International Publishing.

46. Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2010). Individual differences in gameplay and learning: a narrative-centered learning perspective. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games* (pp. 171-178). ACM.

47. Shelton, B. E., & Scoresby, J. (2011). Aligning game activity with educational goals: Following a constrained design approach to instructional computer games. *Educational Technology Research and Development*, *59*(1), 113-138.

48. Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153-189.

49. Shute, V. J., Rieber, L., & Van Eck, R. (2011). Games... and... learning. *Trends and issues in instructional design and technology*, 3.

50. Sicart, M. (2008). Defining game mechanics. *Game Studies*, *8*(2), 1-14.

51. Squire, K., Barnett, M., Grant, J. M., & Higginbotham, T. (2004, June). Electromagnetism supercharged!: learning physics with digital simulation games. In *Proceedings of the 6th international conference on Learning sciences* (pp. 513-520). International Society of the Learning Sciences.

52. Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, *120*(3), 439–471.

53. Tawfik, A. A., Moore, J. L., He, Z., and Vo, N. (2012). Human-computer interaction factors in designing educational video games. *Current Issues in Education*, *15*(3), 1-18.

54. Tulving, E., & Donaldson, W. (1972). *Organization of memory*. Academic Press, New York, NY.

55. van Merriënboer, J.J.G., Jelsma, O. and Paas, F.G.W.C. (1992). Training for reflective expertise: A four-component instructional design model for training complex cognitive skills, *Educational Technology Research and Development*, 40(2), 23–43.

56. van Merriënboer, J. J. G., & de Croock, M. B. M. (2002). Performance-based ISD: 10 steps to complex learning. *Performance Improvement*, *41*(7), 33–38.

57. Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, *34*, 229–243. doi:10.2190/FLHV-K4WA-WPVQ-H0YM

58. Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

59. Wetzstein, A., & Hacker, W. (2004). Reflective verbalization improves solutions—the effects of question-based reflection in design problem solving. *Applied Cognitive Psychology*, *18*(2), 145-156.

60. Wilkinson, G. L. (1980). Media in instruction: 60 years of research. Publications Office, Association for Educational Communications and Technology, 1126 16th St., NW, Washington, DC 20036.

61. Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., Orvis, K. L., & Conkey, C. (2009). Relationships between game attributes and learning outcomes review and research proposals. *Simulation & gaming*, *40*(2), 217-266.

62. Young, M. F., Slota, S., Cutter, B., R., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research*, *82*(1), 61–89. http://doi.org/10.3102/0034654312436980

63. Zohar, A., & David, A. B. (2008). Explicit teaching of meta-strategic knowledge in authentic classroom situations. *Metacognition and Learning*, *3*(1), 59-82.

11. Appendix A: Survey

Algemene vragenlijst

Geslacht: Man ☐ Vrouw ☐

Leeftijd: .....

Studie: .....

Heb je werkervaring in de verzorging van patiënten in een verzorgingstehuis, ziekenhuis of bij de patiënt thuis?

☐ Nee

☐ Ja, maar minder dan een jaar

☐ Ja, meer dan een jaar

Als je de vorige vraag met ja hebt beantwoordt, heb je ook een opleiding gevolgd om zorgverlener te worden?

☐ Nee

☐ Ja

Hoeveel tijd besteed je gemiddeld per week aan het spelen van digitale spellen?

☐ 0 uur per week

☐ 1 – 3 uur per week

☐ 4 – 6 uur per week

☐ 7 – 10 uur per week

☐ Meer dan 10 uur per week

Indien je digitale spellen speelt, wat voor soort speel je zoal? (je mag meerdere soorten aanvinken)

☐ Adventure (b.v. Uncharted, Zelda)

☐ Role-playing game (b.v. Final Fantasy, The Elder Scrolls 5: Skyrim,)

☐ Sport/Racing (b.v. Fifa, Gran Turismo)

☐ Puzzle (b.v. Portal, Tetris, Candy Crush)

☐ First-Person Shooter (b.v. Call of Duty, Counterstrike)

☐ MOBA (b.v. League of Legends, DotA)

☐ MMORPG (b.v. World of Warcraft, Runescape)

12. Appendix B: Manual

Instructieblad voor “Kun je helpen?”

Lees de instructies goed door, want je mag er na level 3 niet meer naar kijken!

Doel:

Help Eragon bij het uit laten komen van zijn drakenei door hem instructies te geven over hoe hij het ei naar het kampvuur moet verplaatsen.

Hoe speel ik het spel?

Je speelt het spel door Eragon een selectie van acties te geven, waarna je op ‘play’ klikt om hem deze uit te laten voeren. Als het Eragon met jouw selectie niet lukt krijg je feedback en mag je een nieuwe selectie maken.

Stap 1: Klik op een vak in de tijdlijn (deze kan al een geselecteerde actie bevatten).

Stap 2: Klik op een van de 2 categoriën, of verwijder een al geselecteerde actie door op de prullenbak te klikken.

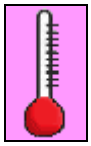
Stap 3: Klik op een van de 5 acties om de actie in het gekozen vak te plaatsen. **Het vraagteken geeft uitleg over de actie.**

Stap 4: De actie is geselecteerd en maakt nu deel uit van je selectie, ga door met het volgende vak.

Belangrijk! Je hoeft niet de complete tijdlijn in te vullen, alle levels zijn met **minder** dan **8** acties te behalen!

Wat zijn de regels?

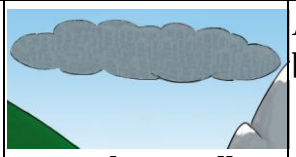
Het drakenei en de wereld waarin de levels zich afspelen kunnen bepaalde kenmerken hebben waaruit je kan afleiden hoe je het ei naar het kampvuur moet verplaatsen. Deze kenmerken kun je vinden in de volgende tabel:

Wat	Kenmerk	Waardes/indicatie	Bijbehorende regel
Ei (zicht-baar)	**Gewicht**	 **50 kg**	Kan alleen met een **lift getild** én **gedragen** worden.
		 **20 kg**	Kan **getild** worden door **Eragon** of **getild** én **gedragen** worden door een **lift**.
		 **15 kg**	Kan getild én gedragen worden door Eragon of door een lift.
Ei (onzicht-baar)	 **Temperatuur**	 **Ziek**	Ziekte kan 2 dingen betekenen: 1. **Het ei is 30 °C (Koel)** **OF** 2. **Het ei is 70 °C (Heet)** - Een heet ei (**70 °C**) vereist **handschoenen** voor optillen en dragen **zonder lift**. - Een koel ei (**30 °C**) kan zonder problemen opgetild en gedragen worden.
	 **Emotie**	 **Slapende**	Het ei kan na wakker worden: 1. **Boos zijn** **OF** 2. **Bang zijn** - Een **bang** ei kan **niet** met de **lift** verplaatst worden. - Een **boos** ei vereist **handschoenen** voor optillen en dragen **zonder lift**.

Onzichtbare kenmerken meten:

Als het ei ziek is of slaapt, betekent het dat er één of meerdere metingen gedaan moeten worden om te kijken of er rekening gehouden moet worden met bepaalde kenmerken. Rechtsbovenin elk levelscherm heb je de mogelijke metingen staan voor dat level.

Belangrijk! Je mag in geen enkel level alle metingen doen! Je zult na moeten gaan a.d.h.v. alle andere kenmerken welke meting je nodig hebt!

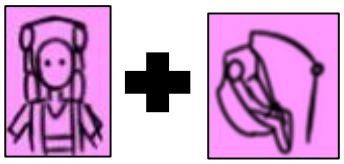
Om-geving	**Terrein**	 **Modder**	**Alleen Eragon** zelf kan over **modder** heenlopen.
		 **Grind**	Alleen de **rolstoel** of **Eragon** zelf kunnen over **grind** heen.
		 **Gras**	De **lift**, zowel als de **rolstoel** en **Eragon** zelf kunnen over **gras** heen.
	Weer	 **Zon**	Als de **zon** schijnt is de **temperatuur** van het **ei** altijd **70 °C**.
		 **Donkere wolk**	Als er een **donkere wolk** hangt is het **ei** altijd **bang**.
Extra	**Combinatie**	**N/A**	Als een **ei** zowel **boos** is als **70 °C**, is een **lift** vereist.

Hoe verplaats ik het ei?

Er zijn **3** manieren waarop het ei verplaatst kan worden:

Zelf tillen: Eragon kan voorwerpen **dragen** die maximaal **15 kg** wegen.

Rolstoel: Eragon kan voorwerpen **tillen** die maximaal **23 kg** wegen. Begin altijd pas met tillen als de rolstoel er staat. Plaats het ei op de rolstoel en duw het naar de overkant.



Tilband + Passieve verrijdbare tillift (PVT): Als Eragon het ei niet meer kan tillen, biedt de combinatie van een tilband met PVT uitkomst. Leg de tilband klaar, til het ei met de lift op en duw het naar de overkant.
Deze manier van **verplaatsen** is geschikt om **fysiek contact** te **vermijden.**

Heb je aan de hand van dit instructieblad nog vragen, stel ze gerust!