July 2010

Learning the triage procedure.

Serious gaming based on guided discovery learning versus studying worked examples.

V. van Dijk – 3138852

Educational Sciences, Faculty of Social Sciences, Utrecht University

Supervisors: G. Erkens & E. D. van der Spek

Abstract

Background. An increasing number of researchers and educators is considering the use of serious games. Serious games often reflect unguided discovery learning with the assumption that the best learning occurs when students discover and construct information by themselves. However, cognitive load theorists argue that unguided discovery learning generate a heavy working memory load, especially for novices, which is detrimental to learning. A possible way to reduce cognitive load is using worked examples. **Aim.** The aim of this study was to explore what the best way to learn the triage procedure is (categorizing victims of a large-scaled accident according to urgency of needed medical attention) by comparing a serious game based on guided discovery learning with a PowerPointpresentation based on worked examples. **Sample**. The participants were 49 students (mean age of 21.15 years) of different studies from higher vocational education or university. **Method**. The design was pre-, post-, delayed post-test with an assessment of constructed mental model; measurements of retention and application; measurements of experienced cognitive load during and immediately after the acquisition phase and measurement of motivation immediately after the acquisition phase. **Results**. In the short term there was a statistically significant difference on the total, retention and application post-test in favour of the worked example condition. No statistically significant difference was found on the assessment of the constructed mental model. In the long term no statistically significant differences were found. However, the guided discovery condition declined statistically significantly less on the total and retention test. The guided discovery condition experienced statistically significantly more cognitive load. No difference in motivation was found. Where there was a statistically significantly negative correlation between averaged experienced cognitive load and (delayed) post-tests and a statistically significantly positive correlation between motivation and (delayed) post-tests in the guided discovery condition, no correlations were found in the worked example condition. In both conditions, there was a statistically significantly negative correlation between averaged experienced cognitive load and motivation. **Conclusion**. With respect to long term learning, serious games based on guided discovery learning seems somewhat promising. However, limitations and future recommendations are discussed.

Keywords: Cognitive load theory; Discovery learning; Serious games; Worked examples

1. Introduction

A new generation of students, who is digitally literated and raised in the presence of video, console and computer games, is entering education (Oblinger, 2004). An increasing number of researchers and educators is considering the use of games as a means of teaching this new generation, for at least three reasons. First, there has been a major shift in the field of learning and instruction from a more traditional, didactic model of instruction to a more student-centred model that emphasizes a more active role of the student, which can be provided by an interactive game. Second, there is some empirical evidence that games can be effective tools for enhancing learning and understanding of complex subject matter. However, as will be indicated later, this evidence is not so straightforward. Third, games can involve and engage students and educators will use these motivational aspects of games to enhance learning and accomplish instructional objectives (Garris, Ahlers & Driskell, 2002).

Games intended for learning and instruction are called serious games. In this study a serious game is defined as a learning environment in which the student is engaged in a goal-driven, competitive activity within a framework of agreed rules. This competition can be against the computer, another student or oneself. While playing the game the student constantly receives feedback enabling him to monitor his progress towards the goals he has set. The main difference between a game and a simulation is the competitive factor (Wouters, Van der Spek & Van Oostendorp, 2008).

One kind of a serious game is an adventure game. An adventure game is a genre of role-playing games in which the player assumes the role of a character in a fantasy world. The player is able to move around in this world, inspect it and interact with whatever is found. The fantasy world is itself rarely static, since other characters and objects can move about and act on their own. Furthermore, an adventure game contains challenges, such as puzzles to be solved or discovering hidden contingencies in the fantasy world. In addition, an important aspect of an adventure game is the context and the associated set of tasks that must be completed if the player is to advance in the game (Halff, 2005). It should be noted that it is possible that the game world of a serious game reflects a more realistic world instead of a fantasy world.

1.1. Serious Games and Motivation

According to Clark, Yates, Early and Moulton (in press) a constant focus of attention in learning and instruction is the exciting developments in new technologies and media. The most common assumption is that new media are more motivating than older media and that increased motivation will lead to significantly more learning and performance. Serious games can also be seen as an exciting development and one of the main reasons to use serious games in education is the assumed motivation for students. According to Prensky (2003) motivation plays an important role in successful learning: “A sine qua non of succesful learning is motivation: a motivated learner can’t be stopped” (Prensky, 2003, p. 1). However, in their review of the effectiveness of serious games Wouters, Van der Spek and Van Oostendorp (2009) found no compelling evidence for the assumed motivation of serious games. Given the popularity of playing games among the new generation of students, it seems clear that games are motivating, but it is unclear to what extent this pertains to serious games. This is also found in the review of Hays (2005) who said that the claim that games improve the motivation and interest of students is still weakly supported.

However, Randel, Morris, Wetzel and Whitehill (1992) found in their review that in 12 out of 14 studies students reported more interest in simulations/games activities than in more conventional classroom instruction.

1.2. Serious Games and Learning

Little empirical research has been done to investigate the learning outcomes of serious games, and contrary findings have been found. Based on a review of 26 review studies and 48 empirical studies on the effectiveness of serious games, Hays (2005) concluded that the empirical research on the effectiveness of serious games is fragmented and included research on different tasks, participants and games. Furthermore, the research literature is filled with ill-defined terms and methodological flaws. He also concluded that serious games could enhance cognitive learning; people could learn from serious games. However, the research did not indicate that serious games were superior to other instructional methods in all cases.

In an earlier review based on 68 articles between 1984 and 1991 that compared the instructional effectiveness of games with conventional classroom instruction, Randel and colleagues (1992) found that 38 studies showed no differences between games and conventional classroom instruction; 27 favoured games, but in 5 studies the controls were questionable; and 3 favoured conventional classroom instruction. Furthermore, simulations/games showed greater retention over time than conventional classroom instruction.

Based on a recent review of 28 studies between 2001 and 2008 with empirical data about the learning outcomes of serious games, Wouters and colleagues (2009) concluded that serious games seemed to be effective for cognitive learning outcomes; the use of serious games for training motor skills and attitudinal change was promising; and little recent substantiation is found for the effectiveness of serious games on communicative learning outcomes.

2. Serious Games and (Un)guided Discovery Learning

Although serious games can reflect different learning theories, many employ unguided, discovery, constructivist or problem-based learning theories (Clark et al., in press). Modern serious game designers believe that the primary strength of serious games is the ability to present students with authentic problems in realistic contexts that facilitate inquiry and exploration (Kebritchi & Hirumi, 2008). Serious games allow students to discover new rules and ideas rather than memorizing material presented by others (Kiili, 2005).

Unguided discovery learning requires students to discover or explore concepts, procedures, phenomena and/or problems without any guidance and construct essential information for themselves (Hmelo-Silver, Duncan & Chinn, 2007; Kirschner, Sweller & Clark, 2006; Mayer, 2004; Tuovinen & Sweller, 1999). It takes place in a learning environment where the student receives problems to solve with little or no guidance and cueing from the teacher or other sources (Gagné, 1971; Mayer, 2004). Students are expected to find regularities and relationships in the learning environment. With this expectation, they devise strategies for searching and finding out what these regularities and relationships are (Driscoll, 2005). Learning environments based on unguided discovery learning challenge students to solve authentic problems or acquire complex knowledge in information-rich settings based on the assumptions that when students construct their own solutions this leads to the most effective learning experience and that a greater amount of self-generated processing results in superior memory and performance of the target skill than a more passive encoding method (Kirschner et al., 2006; Reiser, Copen, Ranney, Hamid & Kimberg, 1998). Advocates of unguided discovery learning assumed that the best learning occurs when students discover their own solution to a problem or task; that students are more likely to remember concepts that they discover on their own and that they acquire more elaborated or robust knowledge, which is better understood, better organized and more amenable to future use, than students who merely observe solutions performed by other people or read worked-out solutions (Clark et al., in press; Kebtrichi & Hirumi, 2008; Reiser et al., 1998). It is also assumed that unguided discovery learning facilitates long-term retention and transfer (Driscoll, 2005). Furthermore, it is assumed that students learn by inferring principles rather than being told them and learn from errors, since an important advantage of exploration is the potential for encountering unanticipated events. Students have to learn to reproduce unanticipated positive events and to avoid negative ones (Reiser et al., 1998). Schank, Berman and MacPherson (1999) proposed that students learn when their expectations fail and they attempt to explain the unanticipated success or failure. Therefore a learning environment should encourage students to ask questions, to predict answers, to be wrong and to explain, thereby profiting from errors (Schank et al., 1999).

However, it has been found that unguided discovery learning is often insufficient for adequate learning, since students engage in trial-and-error behaviour and make many wrong moves before figuring out the appropriate sequence of actions necessary to solve a problem. Students are often unable, from exploration alone, to discover the most efficient sequence of actions to attain a given goal in the environment (Merrill, 1999). Moreover, investigations of novices who solve problems, have suggested that errors can often be costly and lead to counterproductive floundering (Reiser et al., 1998). There is also evidence that unguided discovery learning may have negative results when students acquire misconceptions or incomplete and/or disorganized knowledge (Kirschner et al., 2006). Another potential disadvantage of unguided discovery learning is that students fail to discover all the important aspects of the domain (Reiser et al., 1998).

The review studies of Mayer (2004) and Kirschner and colleagues (2006) provided convincing evidence that unguided discovery learning is almost always less effective than guided discovery learning. In a learning environment based on guided discovery learning, the student also receives problems to solve, but the teacher, or other sources, also provide hints, direction, coaching, feedback, modeling and/or cueing (Gagné, 1971; Mayer, 2004). The advantage of guided discovery learning over unguided discovery learning is especially the case for novices, who lack proper schemas to integrate the new information with their prior knowledge (Schmidt, Loyens, Van Gog & Paas, 2007). Furthermore, students with no schemas or very limited schemas associated with the domain of study may find it difficult to decide which aspects of the learning environment to explore, while students with additional schemas through experience find it easier to make these exploration decisions (Tuovinen & Sweller, 1999). In addition, Driscoll (2005) stated that students without prior knowledge will undoubtedly face frustration and failure in an unguided discovery learning environment. One reason for this is that the discovery process, such as searching and finding out what the regularities and relationships are, is not compatible with the human cognitive architecture and requires a huge amount of unproductive mental effort, which is especially the case for novices.

According to Kirschner and colleagues (2006), even for students with considerable prior knowledge, strong guidance during learning is most often found to be equally effective as unguided learning. However, Van Merriënboer and Kirschner (2007) stated that if students have well-developed knowledge and skills, guided discovery instruction builds heavily on these and offers them good opportunities for using elaboration skills. That is, embedding new information in already available cognitive schemas. Guided discovery learning may be usable if there is ample time available; if students have well-developed knowledge and skills; and if a deep level of understanding is necessary (Van Merriënboer & Kirschner, 2007). Furthermore, Kalyuga, Chandler, Sweller and Tuovinen (2001) stated that the guidance can interfere with the already constructed schemas of experts and that offered guidance can become redundant for them.

Summarized, advocates of unguided discovery learning claimed that students learn the best when they discover and construct information by themselves. However, cognitive load theorists claimed that unguided discovery of a highly complex environment may generate a heavy working memory load, especially for novices, which is detrimental to learning. They argue for more guidance in the learning environment.

In addition to the fact that serious games often reflect unguided discovery learning, playing a serious game is a complex task, since the student has to visually attend different locations on the screen; combine this with navigation with a mouse or joystick; interpret different visual, auditory, verbal and/or tactile cues; and solve problems, often within a time limit. In combination with the human cognitive architecture, which is described below, this complex task may cause problems to learning and instruction (Wouters et al., 2008).

3. Human Cognitive Architecture

The human cognitive architecture is characterized by a limited working memory and an unlimited long-term memory. Working memory deals with the conscious processing of information (Sweller, Van Merriënboer & Paas, 1998). However, it is limited in two ways. First, the amount of elements that can be simultaneously held in working memory is limited to 7 ± 2 elements, but the amount of information of one element is not limited. One way to increase the capacity of working memory is thus through the creating of larger information elements, also known as the process of chunking. Second, the amount of time information can be retained in working memory without rehearsal is about 15 to 30 seconds (Driscoll, 2005). Working memory is thus limited in both capacity and duration.

In contrast to working memory, long-term memory is the unconscious part of the human cognitive architecture with an unlimited capacity and a long duration (Gray, 2007). It can hold an unlimited number of information elements in the form of schemas varying in their degree of automaticity. Schemas have two functions. First, they provide a mechanism for the organization and storage of knowledge in long-term memory. Second, they reduce working memory load, since a schema can be treated as a single information element in working memory. Schemas can be consciously or automatically processed in working memory. Automated schemas decrease working memory load further, since conscious processing is not necessary, thereby freeing working memory capacity for other activities (Sweller et al., 1998).

4. Cognitive Load Theory

The cognitive load theory is based on the characteristics of the human cognitive architecture. Schema construction and automation are the primary goals of instruction. In order to construct schemas for long-term memory, information has to be processed in working memory. The ease with which information is processed in working memory is a prime concern of cognitive load theory, since working memory is limited in capacity and duration. In the cognitive load theory, three types of cognitive load can be distinguished.

First, intrinsic cognitive load is a direct function of the structure and complexity of the learning material. The working memory load imposed by the learning material depends on the number of elements that must be simultaneously processed in working memory and this is dependent on the extent of element interactivity. The element interactivity is low when elements of the learning material can be learned in isolation; and high when elements of the learning material have to be simultaneously manipulated in working memory. The higher the element interactivity, the higher the intrinsic cognitive load. It is important to note that the element interactivity can not be solely determined by analysing the learning material, since it depends also on the expertise of the student. Learning material can have a high element interactivity for some students, especially novices, since they have no automated constructed schemas in long-term memory. The result is a high cognitive load of working memory. However, the same learning material can have a low element interactivity for experts, since they have automated constructed schemas in long-term memory which can be used as single information elements in working memory, resulting in a low cognitive load of working memory. Since the element interactivity is intrinsic to the learning material, it can not be altered by instructional designers.

Second, extraneous cognitive load is caused entirely by the design of the learning material and does not contribute to learning. Therefore it is the task of instructional designers to decrease this extraneous cognitive load, especially when the intrinsic cognitive load is high, since extraneous and intrinsic cognitive load are additive.

Third, germane cognitive load contributes to learning, or in other words, to schema construction. It can be altered by instructional designers, so it is their task to increase the germane cognitive load in order to increase learning. Underlying idea is that an instructional design that results in unused working memory capacity because of low intrinsic cognitive load, low extraneous cognitive load, or a combination of both, may be further improved by encouraging students to engage in conscious cognitive processing in order to increase learning. However, increasing germane cognitive load is only possible when either intrinsic or extraneous cognitive load is low, or when a combination of both applies. This is because the total cognitive load, the sum of intrinsic, extraneous and germane cognitive load, has to stay within the limitations of working memory (Sweller et al., 1998).

5. Worked Examples

Cognitive load theorists claim that unguided discovery learning is not compatible with the human cognitive architecture, especially for novices, and is therefore detrimental for learning. According to them, it is the task of instructional designers to decrease the extraneous cognitive load, caused by the discovery character of the learning environment, and to increase the germane cognitive load in order to increase learning.

In the domain of problem solving one way to decrease the extraneous cognitive load is using worked examples instead of conventional problem solving. A worked example consists of a problem statement followed by all the appropriate steps necessary to reach the solution. Underlying idea is that by studying worked examples the student did not have to allocate cognitive resources to the problem solving process itself, but can allocate all his cognitive resources to the studying of the appropriate steps necessary to reach the solution in order to increase the potential of schema construction and automation. The advantage of worked examples over conventional problem solving is called the worked example effect (Kalyuga, Chandler & Sweller, 2001; Kalyuga et al., 2001; Sweller et al., 1998).

The worked example effect is obtained in different domains, such as algebra (Carroll, 1994; Cooper & Sweller, 1987; Sweller & Cooper, 1985;), statistics (Paas, 1992), geometrical problems (Paas & Van Merriënboer, 1994), databases (Tuovinen & Sweller, 1999), writing simple programmable logic controller programs and Boolean switching equations for relay circuits (Kalyuga et al., 2001; Kalyuga et al., 2001), diameter circumference conversion charts (Kalyuga et al., 2001), physics (Lee, Nicoll & Brooks, 2004) and geometry (Schwonke, Renkl, Krieg, Wittwer, Aleven & Salden, 2009).

In general, it can be concluded that studying worked examples is superior to conventional problem solving. However, a few exceptions can be indicated.

First, the worked example effect did not apply when the learning material is simple, or in other words, has a low element interactivity, since this results in a low intrinsic cognitive load whereby the total cognitive load probably stays within the limitations of working memory (Cooper & Sweller, 1987; Kalyuga et al., 2001; Kalyuga et al., 2001; Sweller & Cooper, 1985).

Second, the worked example effect did not apply when students have high prior knowledge and/or experience, since the element interactivity of the learning material is low for them, resulting in low intrinsic cognitive load. This is called the expertise reversal effect. Furthermore, worked examples could become redundant for experienced students, which is called the redundancy effect (Kalyuga et al., 2001; Kalyuga et al., 2001; Tuovinen & Sweller, 1999).

Third, the application of the worked example effect depended on the design of the worked examples. When worked examples include two or more mutually referring sources of information that can not be understood in isolation and have to be mentally integrated, there is no reason to predict that the worked example effect will apply. In this case, students have to split their attention between multiple sources of information and have to allocate their cognitive resources to mentally integrate these different sources, resulting in extraneous cognitive load. This is called the split-attention effect (Kalyuga et al., 2001; Sweller, 2005).

6. The Present Study

There is a discussion about the amount of guidance in the learning environment. On the one hand there are theorists who argue that students learn the best when they discover and construct information by themselves in a less or more guided learning environment. However, on the other hand, cognitive load theorists argue for more guidance in the learning environment, especially for novices. One form of guidance is using worked examples.

The aim of this study was to explore what the best way to learn the triage procedure with a computer based instruction is, by comparing a serious game based on guided discovery learning with a PowerPointpresentation based on worked examples.

A procedure can be defined as: an ordered sequence of steps necessary for the student to attain some goal, solve a particular class of problem or produce some product (Merrill, 1994). Medical first responders use the triage procedure to categorize the victims of a large-scaled accident, according to urgency of needed medical attention. In the Netherlands, and most other Western European countries, this procedure approximately consists of the following four central checks. First, check if the victim can walk; second, check if the victim's airway is free; third, check if the respiratory frequency is within a certain range; and fourth, check if the capillary refill time is within a certain range. When the medical first responder has the results of these checks, he has to categorize the victim into one of four triage categories: T1 for urgent, T2 for delayed, T3 for light or no injury and Dead for death immanent or already deceased (Van der Spek, Wouters & Van Oostendorp, 2009). In addition, if the airway is not free, either a chin lift or jaw thrust have to be used and depending on the outcome, the victim has possible to be placed in the stable recovery position. Furthermore, if it is too dark to check the capillary refill time, the pulse rate has to be checked (Van der Spek, Van Oostendorp, Wouters & Aarnoudse, 2010). The triage procedure is shown in the flowchart in Figure 1.

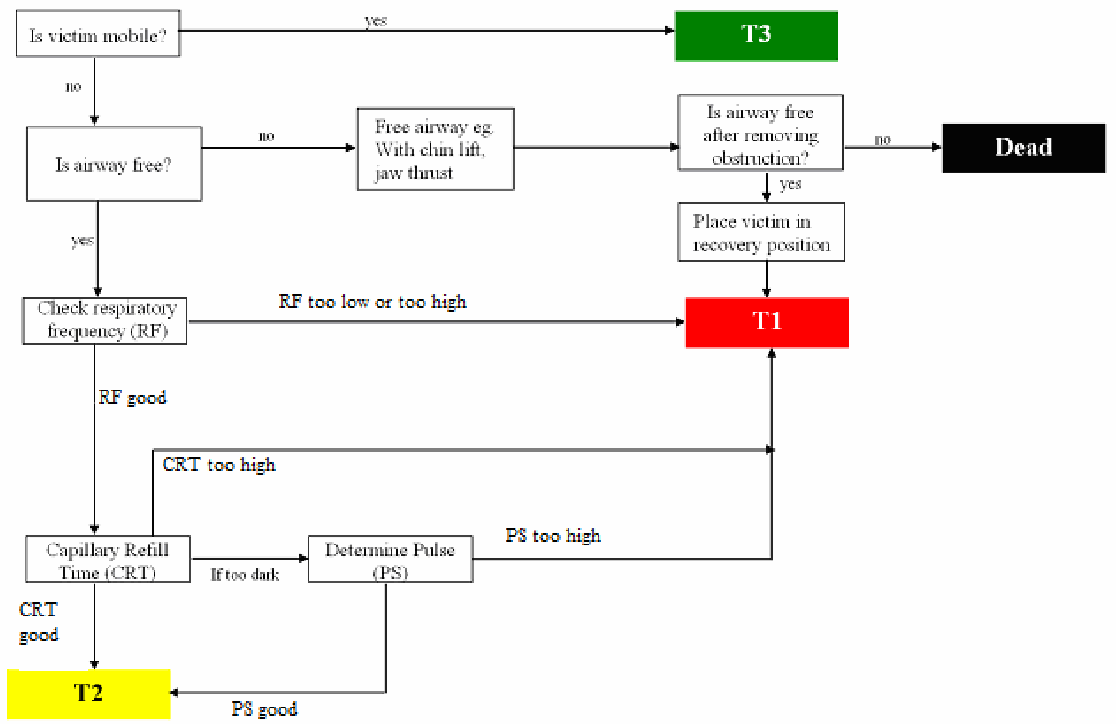


Figure 1. Flowchart of the triage procedure (adapted from Van der Spek et al., 2010).

6.1. Central Research Question and Sub Questions

The central research question is: Is there a difference in learning outcomes for students who played a serious game based on guided discovery learning compared with students who studied worked examples in a PowerPointpresentation, when learning the triage procedure?

To answer this central research question, there are seven sub questions:

1. Is there a learning result after following the instruction?
2. Is there a difference in learning outcomes on the total, retention and application post-tests and in constructed mental model, immediately after the acquisition phase, between the guided discovery and worked example condition?
3. Is there a difference in learning outcomes on the total, retention and application delayed post-tests and in constructed mental model, a week after the acquisition phase, between the guided discovery and worked example condition?
4. Is there a difference in experienced cognitive load during, and immediately after, the acquisition phase between the guided discovery and worked example condition?
5. Is there a difference in motivation, immediately after the acquisition phase, between the guided discovery and worked example condition?
6. Can the learning outcomes be explained by averaged experienced cognitive load?
7. Can the learning outcomes be explained by motivation?

7. Method

7.1. Participants and Design

The participants were 49 students, from different studies in higher vocational education (8 participants) or university (41 participants) in the Netherlands. It was assumed that the participants were not familiar with the triage procedure. This assumption was tested with a pre-test (see section 7.3.). Due to technical problems, one participant of the guided discovery condition was removed.

Twentyfour students participated in the guided discovery condition, 6 male and 18 female with a mean age of 21.08 years (SD = 1.47). Five students were from higher vocational education and 19 from university. Twentyfour students participated in the worked example condition, 6 male and 18 female with a mean age of 21.21 years (SD = 2.23). Three students were from higher vocational education and 21 from university.

With respect to the delayed post-tests, one participant of the worked example condition did not send back the delayed post-tests. Furthermore, one participant of the guided discovery condition sent back the delayed post-tests far too late and was therefore removed from the analysis.

The design was experimental with two independent groups, since the participants were randomly assigned to the two conditions. A pre-, post-, delayed post-test design was used.

7.2. Interventions

7.2.1. Serious game based on guided discovery learning.

The serious game was an interactive, 3D, first person computergame, called Code Red: Triage, which could be classified as an adventure game, since the student assumed the role of a character (first medical responder); there was a context (a terrorist strike on the subway); and there were tasks to perform (classifying victims according to the triage procedure).

The game was created with the Source SDK, a development kit created and distributed freely by VALVe Software alongside The Orange Box (2007), for the creation of modifications of Half Life 2 and other games that run on the Source engine.

In the game Code Red: Triage the student had to discover the appropriate steps of the triage procedure. The game started with a description of the environmental context, namely a terrorist strike on the subway. After this description, the student arrived in the train station and had to find the way to subway platform number two. This part of the game was intended to practice with the navigation of the game without a time limit. A screenshot of the train station is shown in Figure 2.

Figure 2. Screenshot of the train station.

Arriving at the subway platform, the student encountered a total of 19 victim cases with varying degrees of injuries. When the student followed the right route in the game, the victim cases had an increasing difficulty. A screenshot of the subway platform is shown in Figure 3.

Figure 3. Screenshot of the subway platform.

When the student arrived at a victim and pressed the e-button the triage procedure interface menu appeared, which is shown in Figure 4.

Figure 4. Screenshot of the interface menu.

This interface menu consisted of three parts:

1. Eight check buttons represented the different checks of the triage procedure: mobility, airway, respiratory frequency, capillary refill time, chin lift, jaw thrust, stable recovery position and pulse rate. The order in which the check buttons were placed was different from the right sequence of the triage procedure. The number of check buttons gradually increased from two check buttons by the first victim to eight check buttons by the last victim. This was based on the idea of part-whole sequencing of Van Merriënboer and Kirschner (2007).
2. The information of each check button consisted of two parts, namely general information about the specific step of the triage procedure, and the answer of the victim. These parts were divided by a double line space, so the student had the opportunity to skip the general information, which was always the same, thereby reducing redundancy. In some cases, the student had to wait for the

answer of the victim, since some checks cost time to perform, such as the measurement of the pulse rate.

3. The bottom of the interface menu consisted of four buttons to classify the victim into one of four triage categories (T1 = red, T2 = yellow, T3 = green or Dead = black). After classifying the victim, the victim was coloured according to the colour of the triage category.

Guidance was given by placing the victims in an easy-to-difficult sequence and by gradually increasing the amount of check buttons. Furthermore, the student received guidance in the form of feedback and a score. The feedback consisted of four parts, namely if the student:

1. performed the triage within the prescribed time;
2. did the steps of the procedure in the right sequence;
3. forgot steps;
4. did unnecessary steps.

Positive feedback was coloured green and negative feedback was coloured red. The maximum score for the correct classification of a victim was 100 points, with a maximum of 1900 points for the whole game. This score was shown as a progress bar at the top of the game. The student received penalties, if he...

1. exceeded the prescribed time;
2. did steps of the procedure in the wrong sequence;
3. forgot necessary steps and/or;
4. did unnecessary steps.

When the student again classified a victim, he received the following feedback: Victim already classified.

The game was self-paced with a time limit of 17 minutes and was automatically closed when the time ran out or when the student had classified all the 19 victims.

7.2.2. PowerPointpresentation based on worked examples.

The PowerPointpresentation was made with Microsoft PowerPoint and consisted of 23 slides: one title slide, one slide with a description of the environmental context, 19 worked examples, one slide to indicate that the student had to answer a paper-and-pencil question and one slide to indicate the end of the PowerPointpresentation. Since the instruction in the guided discovery condition was self-paced, the PowerPointpresentation was also self-paced with a maximum of 14 minutes for the whole PowerPointpresentation. The difference in maximum time between the two conditions was due to the fact that students in the guided discovery condition needed time to navigate to the next victim and sometimes had to wait for a measurement.

The PowerPointpresentation started with a description of the environmental context of the worked examples, namely a terrorist strike on the subway. This description was followed by 19 worked examples, called cases, in an easy-to-difficult sequence. This meant an increasing amount of steps necessary to classify the victim. As mentioned earlier, a worked example consisted of a problem statement followed by all the appropriate steps necessary to reach the solution. This resulted in the structure of the worked examples as shown in Table 1.

Table 1. *Schematic outline of each worked example (case) and an example.*

Worked example	General	Example
Problem statement.	Introduction information: the age of the victim and, if applicable, the injuries.	Adult man of 35 years.
All the appropriate steps.	All the appropriate steps of the triage procedure, where each step consisted of two parts: general information about the specific step of the triage procedure and information dependent on the victim.	First, you check the mobility. ***Hereby you determine if the victim can independently leave the disaster area. If the victim can walk, this means that he or she is lightly injured and you can classify him or her in a triage category.*** The victim says that he can walk, but he will get away. He asks where he has to go.
Solution.	Solution of the problem: You classify the victim in triage category ...	You classify the victim in triage category T3.

Beside the worked example, each slide consisted of a static picture of the victim, which was a screenshot from the instruction of the guided discovery condition. The text was placed at the left and the static picture at the right side of each slide.

As discussed earlier, the worked example effect depended on the design of the worked examples (Kalyuga et al., 2001; Kalyuga et al., 2001; Sweller, 2005; Tuovinen & Sweller, 1999). Therefore the split-attention and redundancy effect were taken into account when designing the worked examples. The split-attention effect was tried to reduce in two ways. First, the text of each case was placed on one slide, thereby reducing the students' need to split attention between two or more slides. Second, the static picture of the victim at the right was placed close to the text at the left side of each slide.

As can be seen in Table 1, the explanation of the general information about the specific step of the triage procedure was written in italic and bold. The choice for this was deliberate, since this explanation was repeated in every slide and could become redundant for the student. By writing the explanations in italic and bold, the student had the opportunity to skip these explanations, thereby reducing the redundancy effect.

An example of a slide is shown in Figure 5.

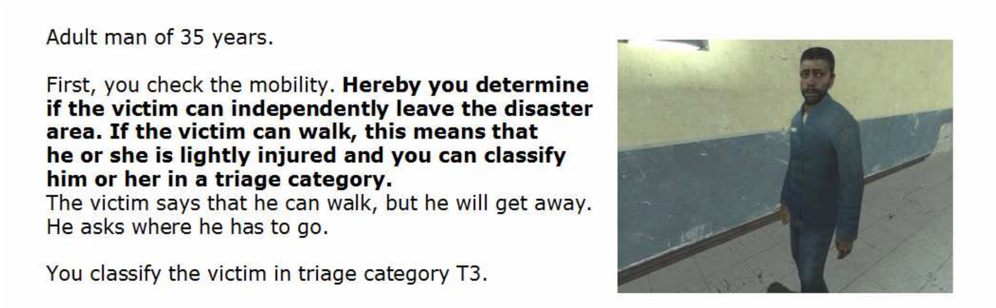


Figure 5. Example of a slide.

The content of both instructions was based on information from three Dutch content experts who work respectively at the *GHOR (Geneeskundige hulpverlening bij ongevallen en rampen)*, the *UMC (Utrechts Medisch Centrum)* and the army. The Dutch book *Geneeskundig management bij grootschalige incidenten* (Hustinx, Meeuwis & Hermans, 2004) was also used.

7.2.3. Explanation about the instruction.

For both conditions a paper-based explanation about the instruction was written. Both explanations informed the student about the aim of the triage procedure (to classify the victim into one of four triage categories); the meaning of the four triage categories (T3: not urgent; T2: quite urgent; T1: highly urgent; and Dead); and the aim of the instruction (to learn the triage procedure as quickly and correctly as possible). In both explanations it was stressed that the sequence of procedure steps was always the same and that the procedure always started with the same step. The student was also informed about the available time to complete the instruction and that he had to answer a paper-and-pencil-based question during the instruction.

Furthermore, in the explanation of the guided discovery condition, information was given about the navigation in the game, the interface menu and the scoring. The whole explanation of the guided discovery condition can be found in Appendix A.

In addition, the explanation of the worked example condition informed the student about the structure of the whole PowerPointpresentation and each case. The whole explanation of the worked example condition can be found in Appendix B.

7.3. Measurements

7.3.1. Retention pre-test, post-test and delayed post-test.

The retention test was used to measure the performance category remembering of the triage procedure. This performance category could be defined as: a performance requiring the student to search in long-term memory in order to reproduce or recognize some item of information previously known (Merrill, 1994).

The retention test consisted of ten if-then, multiple-choice, questions divided in five textual and five visual questions. This division had been made, since the instruction in the worked example condition was highly textual, while the instruction in the guided discovery condition was highly visual. Each question consisted of the if-part of the rule and the six answer options consisted of possible then-parts of the rule. An example of a textual retention question is: If the victim's airway is free, then...

A. you have to check the respiratory frequency.
B. you classify the victim in triage category T3.
C. you have to check the mobility.
D. you have to check the pulse rate.
E. you have to use the jaw thrust.
F. you have to check the capillary refill time.

The right answer is A.

An example of a visual retention question is shown in Figure 6. The right answer is T3.

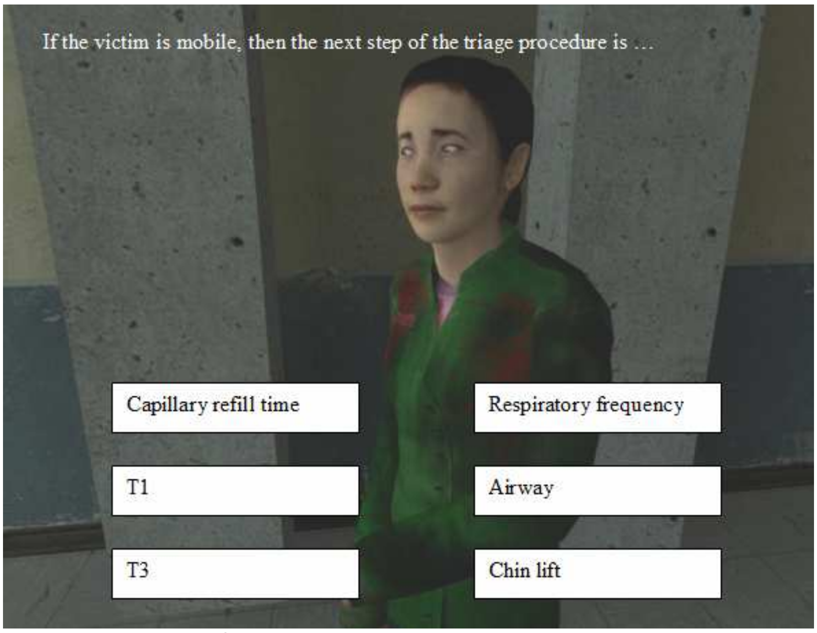


Figure 6. Example of a visual retention question.

In contrast to the answer options of the textual questions, the answer options of the visual questions were just keywords of the triage procedure. The reason for this was that keywords were also used in the instruction of the guided discovery condition.

The three retention tests were the same, only the sequence of the questions was different. For the whole retention test see Appendix C.

7.3.2. Application pre-test, post-test and delayed post-test.

The application test was used to measure the performance category using the triage procedure. This performance category could be defined as: a performance requiring the student to apply some abstraction to a specific case (Merrill, 1994). The application test consisted of ten multiple-choice questions, divided in five textual and five visual questions, with six answer options. Each question consisted of a victim case and the question which action you have to do following the triage procedure in this situation. Each answer option consisted of a possible step of the triage procedure.

When writing the cases it was important that the student could not find the answer of one question in another case. To prevent this, some steps of the triage procedure were switched or removed from some cases.

An example of a textual application question is: A woman of about 20 years with injuries to the torso. The victim can not walk, but does have a free airway. Which action do you have to do following the triage procedure in this situation?

A. Check the pulse rate.
B. Check the capillary refill time.
C. Classify the victim in triage category T1.
D. Check the respiratory frequency.
E. Use the chin lift.
F. Classify the victim in triage category T2.

The right answer is D.

An example of a visual application question is shown in Figure 7. The right answer is T1.

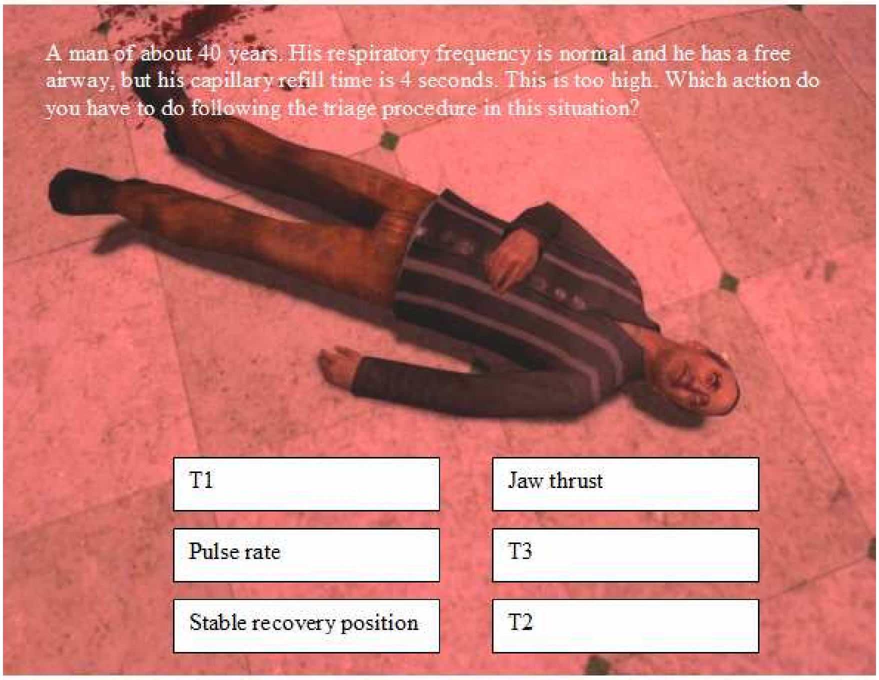


Figure 7. Example of a visual application question.

The three application tests were the same, only the sequence of the questions was different. For the whole application test see Appendix D.

7.3.3. Pre-, post- and delayed post-assessment of the constructed mental model.

A mental model could be defined as the students' cognitive representation of a task and is developed through experience. With the constructed mental model strategic knowledge could be assessed. Strategic knowledge is the combination of declarative and procedural knowledge. It is compiled, automatic and tacit, which means difficult to verbalize, due to continued rehearsal of routine activities (Kahler, 2001). By combining declarative and procedural knowledge in a mental model the student not only knows what (declarative) and how (procedural) to do, but also why and when to do this. A mental model could be transferred between situations; it could be generalize to other situations and used to solve new problems (Van der Spek et al., 2009). Strategic knowledge is also known as conditional knowledge, which enables students to determine when and how to apply declarative or procedural knowledge (Driscoll, 2005).

In order to represent and assess the constructed mental model of the students the structural knowledge method Pathfinder Scaling Algorithm was used. By means of the Pathfinder Scaling Algorithm, which was implemented using the PCKNOT software, a network could be built of the concepts of the task domain under consideration and this was called the Pathfinder Network. In this network, nodes represented the concepts and links represented the relationships between the concepts, where closely related concepts were placed close to each other, and less related concepts further away, or with more intermediate nodes (Lau & Yuen, 2010).

To construct the network, participants of the present study were asked to give ratings from 1 (unrelated) to 9 (related) for 28 pairs of items. The items consisted of two conceptual items (lightly injured, classify victim) and six procedural items (mobility, T3, T1, recovery position, chin lift, obstructed airway) to assess what Kahler (2001) called strategic knowledge.

An example of a rating is shown in Figure 8. The two concepts are highly related, so the student has to give an 8 or a 9.

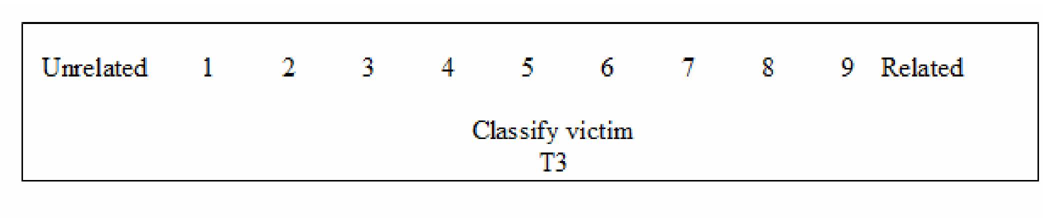


Figure 8. Example of a rating.

The choice to use only a subset of the total triage procedure was intentional, since in earlier tests (cf. Van der Spek et al., 2009) it was found that a larger portion of the procedure was tiresome for the participants and therefore resulted in an unreliable assessment of the constructed mental model.

In order to compare Pathfinder Networks quantitatively, similarity measures can be used by comparing the students' network with a referent mental model of an expert. When corrected for chance, the similarity measure ranges between -1 and 1 (Lau & Yuen, 2010). In order to construct the referent mental model, three researchers independently assessed the 28 pairs and the three obtained networks were averaged to arrive at the referent mental model, which is shown in Figure 9.

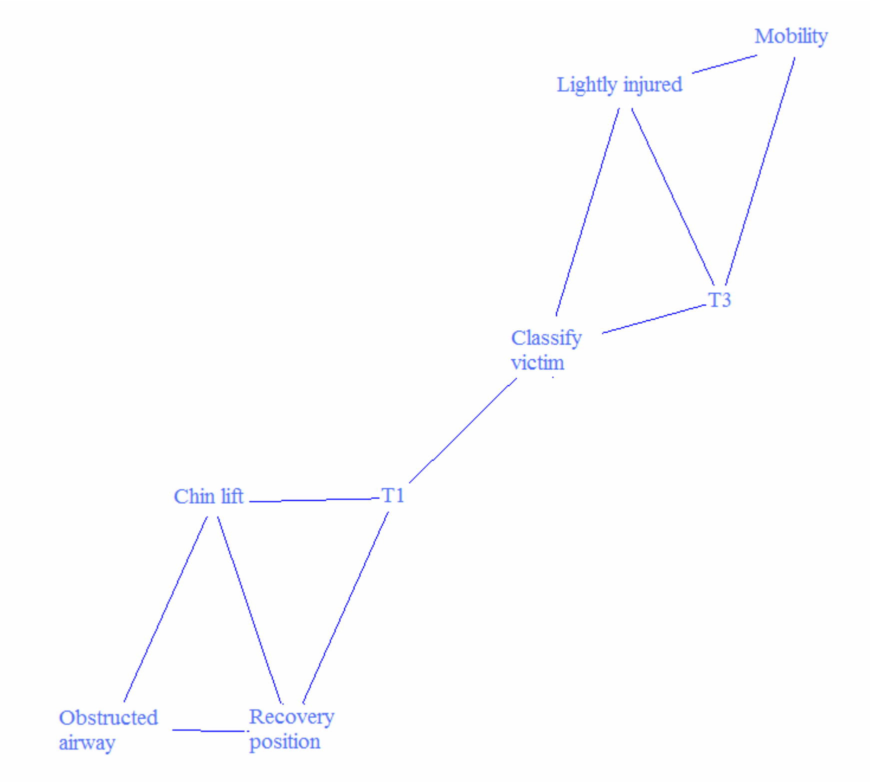


Figure 9. Referent mental model.

The choice to not use practical experts to construct the referent mental model was intentional, since in earlier research (cf. Van der Spek et al., 2009) a number of problems were experienced. For example, experts reported difficulty in looking at the items purely from the perspective of a triage, since they had extensive field experience in other domains too. Furthermore, there were differences between the theoretical triage procedure and the procedure used in the practical field.

The three assessments of the constructed mental model were the same, only the sequence of the pairs was different. For the whole assessment of the constructed mental model see Appendix E.

7.3.4. Measurement of cognitive load.

Since this study was placed in the cognitive load theory, the experienced cognitive load of the participants was measured. There are three major categories of measurement techniques to measure cognitive load, namely subjective, physiological and task- and performance-based indices (Sweller et al., 1998). Subjective measurement of cognitive load was chosen, since this measurement technique was easy to implement and did not intrude on primary task performance (Kalyuga et al., 2001). This was important, since the learning outcomes were also subject of the present study. The primary learning outcomes could also function as a measurement of cognitive load. However, it is possible that participants reach the same learning outcomes with different experience of cognitive load (De Waard, 1996).

The subjective measurement of cognitive load is based on the assumption that participants are able to introspect on their cognitive processes and to report the amount of mental effort expended (Sweller et al., 1998). A 7-point rating scale was used with participants being asked to rate how easy or difficult they found it to understand the instruction. Hereby, a 1 represented “very, very easy” and a 7 represented “very, very difficult”. This question was adopted from Kalyuga and colleagues (2001) and translated into Dutch.

7.3.5. Measurement of motivation.

Since one of the main reasons to use serious games in education is the assumed motivation for students, the participants were asked how entertaining they found the instruction. Like the subjective measurement of cognitive load, a 7-point rating scale was used with participants in the worked example condition being asked to rate how entertaining they found it to read the PowerPointpresentation. Participants in the guided discovery condition were asked how entertaining they found it to play the game. Hereby, a 1 represented “very, very stupid” and a 7 represented “very, very fun”.

It should be noted that the learning outcomes and the measurement of cognitive load were the primary concerns of this study. Therefore only one question was devoted to measure the motivation of the participants.

7.4. Procedure

The experiment took place between 25 March 2010 and 6 May 2010 and was approximately one hour for each participant. The participant received a gift coupon of € 7,50 for participation. The first part of the experiment took place at the Usability Lab of Utrecht University where the participant sat in an enclosed space behind the laptop to ensure a minimum of environmental distractions. The second part of the experiment could be done at home, since the delayed post-tests were e-mailed to the participant. A schematic outline of the procedure is shown in Table 2.

Table 2. *Schematic outline of the procedure.*

First part – Day 1 (t0)	First part – Day 1	First part – Day 1 (t1)	Waiting one week	Second part – Day 8 (t2)
▪ Demographic questionnaire ▪ Pre-test assessment of constructed mental model (5 minutes) ▪ Retention pre-test (5 minutes) ▪ Application pre-test (10 minutes)	▪ Explanation about the instruction ▪ Classifying nine victims ▪ First measurement of cognitive load ▪ Classifying ten victims	▪ Measurement of motivation ▪ Second measurement of cognitive load ▪ Post-assessment of constructed mental model (5 minutes) ▪ Retention post-test (5 minutes) ▪ Application post-test (10 minutes)	-	▪ Delayed post-assessment of constructed mental model (5 minutes) ▪ Retention delayed post-test (5 minutes) ▪ Application delayed post-test (10 minutes)

7.5. Measures and Analyses

The dependent variables under analysis were total, retention and application scores in the short term (t1) and the long term (t2); assessment of the constructed mental model at t1 and t2; experienced cognitive load during and immediately after the acquisition phase; and motivation immediately after the acquisition phase. The independent variable was design of instruction: guided discovery (serious game) or worked examples (PowerPointpresentation).

Since the sample size was not large enough to assume a normal distribution, the Kolmogorov-Smirnov test on the dependent variables was used to test if there was a normal distribution. Since multiple dependent variables were not normal distributed it was decided to use the Mann-Whitney test to test if there was a difference between the two conditions. Furthermore, the Wilcoxon signed-ranks test was used to test if there was a difference between t0, t1 and t2 within the conditions. The Spearman's rankcorrelation was used to test if there was a correlation between averaged experienced cognitive load / motivation and the (delayed) post-tests. For all statistical tests a significance level of .05 was applied and a *p*-value between .05 and .10 was called a trend.

8. Results

In this section, first the reliability of the used tests is presented and second the sub questions are answered in a chronological order.

8.1. Reliability of the Tests

8.1.1. Pre-tests.

The retention pre-test had three outliers. When these outliers were removed, Cronbach's alpha was -.30. The application pre-test had two outliers. When these outliers were removed, Cronbach's alpha was -.13. For both pre-tests there were no options to reach an acceptable level of internal consistency. When both pre-tests were taken together, Cronbach's alpha reached a value of .23, which was not high enough either. A possible explanation for these low internal consistencies is that the students had little prior knowledge about the triage procedure, whereby it was not possible to measure some kind of knowledge construct. Since this is an acceptable explanation, the pre-tests were used in further analyses.

8.1.2. Post-tests.

The Cronbach's alpha of the retention post-test was .74, which is acceptable. The application post-test had a Cronbach's alpha of .54, which is rather low. When question three was removed, Cronbach's alpha reached a value of .59, which is almost acceptable in educational sciences. However, there was no substantive reason to remove this question.

To explore if there was a difference in learning outcomes for remembering and using of the triage procedure, it was decided to use the whole application post-test with the awareness that the internal consistency was rather low.

When both post-tests were taken together, Cronbach's alpha reached a value of .80, which is acceptable.

8.1.3. Delayed post-tests.

The Cronbach's alpha's of the delayed post-test remembering, application and the total delayed post-test were respectively .60, .71 and .81, which are all an acceptable level of internal consistency.

8.2. Is there a Learning Result after Following the Instruction?

Table 3 and Table 4 show respectively the learning results by comparing the post-test (t1) with the pre-test (t0) and the delayed post-test (t2) with the pre-test (t0) for both conditions, by using the Wilcoxon signed-ranks test. For both conditions there were statistically significant differences between all the post-tests and the pre-tests and between all the delayed post-tests and the pre-tests. See Table 3 and 4 for the statistical values.

Table 3. *Learning results comparing t1 with t0 for both conditions.*

		Condition									
		Guided discovery					Worked example				
Variable		n (24)	Mean rank	Sum of ranks	*Z*	*p*	n (24)	Mean rank	Sum of ranks	*Z*	*p*
Total retention post-test – Total retention pre-test	Negative ranks	2	3.00	6.00	-4.12[a]	.00	0	0.00	0.00	-4.20[a]	.00
	Positive ranks	22	13.36	294.00			23	12.00	276.00		
	Ties	0					1				
Retention post-test – Retention pre-test	Negative ranks	2	6.50	13.00	-3.70[a]	.00	0	0.00	0.00	-4.12[a]	.00
	Positive ranks	20	12.00	240.00			22	11.50	253.00		
	Ties	2					2				
Application post-test – Application pre-test	Negative ranks	1	3.50	3.50	-4.01[a]	.00	0	0.00	0.00	-4.22[a]	.00
	Positive ranks	21	11.88	249.50			23	12.00	276.00		
	Ties	2					1				
Post-assessment of constructed mental model – Pre-assessment of constructed mental model	Negative ranks	5	5.80	29.00	-3.46[a]	.00	4	5.50	22.00	-3.53[a]	.00
	Positive ranks	19	14.26	271.00			19	13.37	254.00		
	Ties	0					1				

[a]Based on negative ranks.

Table 4. *Learning results comparing t2 with t0 for both conditions.*

		Condition									
		Guided discovery					Worked example				
Variable		n (23)	Mean rank	Sum of ranks	*Z*	*p*	n (23)	Mean rank	Sum of ranks	*Z*	*p*
Total retention delayed post-test – Total retention pre-test	Negative ranks	0	0.00	0.00	-4.12[a]	.00	1	3.50	3.50	-4.10[a]	.00
	Positive ranks	22	11.50	253.00			22	12.39	272.50		
	Ties	1					0				
Retention delayed post-test – Retention pre-test	Negative ranks	0	0.00	0.00	-4.22[a]	.00	1	2.50	2.50	-4.13[a]	.00
	Positive ranks	23	12.00	276.00			22	12.43	273.50		
	Ties	0					0				
Application delayed post-test – Application pre-test	Negative ranks	1	3.50	3.50	-3.92[a]	.01	1	4.00	4.00	-3.99[a]	.01
	Positive ranks	20	11.38	227.50			21	11.86	249.00		
	Ties	2					1				
Delayed post-assessment of constructed mental model – Pre-assessment of constructed mental model	Negative ranks	7	6.50	45.50	-2.81[a]	.00	6	8.17	49.00	-2.71[a]	.00
	Positive ranks	16	14.41	230.50			17	13.35	227.00		
	Ties	0					0				

[a]Based on negative ranks.

Before answering the remaining sub questions, an overview of all the dependent variables for both conditions is given in Table 5, by using the Mann-Whitney test.

Table 5. *All dependent variables for both conditions.*

	Condition					
	Guided discovery (n = 24)		Worked example (n = 24)			
Variable	Mean rank	Sum of ranks	Mean rank	Sum of ranks	Mann-Whitney U	*p*
Total pre-test	24.19	580.50	24.81	595.50	280.50	.88
Retention pre-test	23.71	569.00	25.29	607.00	269.00	.68
Application pre-test	24.38	585.00	24.62	591.00	285.00	.95
Pre-assessment of constructed mental model	24.94	598.50	24.06	577.50	277.50	.83
Total post-test	19.02	456.50	29.98	719.50	156.50	.01
Retention post-test	19.25	462.00	29.75	714.00	162.00	.01
Application post-test	20.31	487.50	28.69	688.50	187.50	.04
Post-assessment of constructed mental model	25.65	615.50	23.35	560.50	260.50	.57
Total delayed post-test	20.83	479.00	26.17	602.00	203.00	.18
Retention delayed post-test	21.50	494.50	25.50	586.50	218.50	.31
Application delayed post-test	21.22	488.00	25.78	593.00	212.00	.25
Delayed post-assessment of constructed mental model	24.87	572.00	22.13	509.00	233.00	.49
Experienced cognitive load during the acquisition phase	32.94	790.50	16.06	385.50	85.50	.00
Experienced cognitive load immediately after the acquisition phase	30.38	729.00	18.62	447.00	147.00	.00
Motivation immediately after the acquisition phase	26.25	630.00	22.75	546.00	246.00	.37

8.3. Is there a Difference in Learning Outcomes on the Total, Retention and Application Post-Tests and in Constructed Mental Model, Immediately after the Acquisition Phase, between the Guided Discovery and Worked Example Condition?

Before answering this sub question it was tested if there was a statistically significant difference on the pre-tests between the two conditions, since prior knowledge can influence the scores on the (delayed) post-tests. Since there were no statistically significant differences on the pre-tests between the two conditions (see Table 5), it was not necessary to control the scores on the (delayed) post-tests for scores on the pre-tests.

With respect to the learning outcomes at t1, the guided discovery condition scored statistically significantly lower on the total post-test (19.02, 456.50) than the worked example condition (29.98, 719.50), U (24, 24) = 156.50, $p = .01$.

On the retention post-test, the guided discovery condition (19.25, 462.00) scored statistically significantly lower than the worked example condition (29.75, 714.00), U (24, 24) = 162.00, $p = .01$.

On the application post-test, the guided discovery condition (20.31, 487.50) scored statistically significantly lower than the worked example condition (28.69, 688.50), U (24, 24) = 187.50, $p = .04$.

There was no statistically significant difference on the post-assessment of the constructed mental model between the guided discovery condition (25.65, 615.50) and the worked example condition (23.35, 560.50), U (24, 24) = 260.50, $p > .05$.

8.4. Is there a Difference in Learning Outcomes on the Total, Retention and Application Post-Tests and in Constructed Mental Model, a Week after the Acquisition Phase, between the Guided Discovery and Worked Example Condition?

With respect to the learning outcomes at t2, there was no statistically significant difference on the total delayed post-test between the guided discovery condition (20.83, 479.00) and the worked example condition (26.17, 602.00), U (23, 23) = 203.00, $p > .05$.

On the retention delayed post-test, there was no statistically significant difference between the guided discovery condition (21.50, 494.50) and the worked example condition (25.50, 586.50), U (23, 23) = 218.50, $p > .05$.

On the application delayed post-test, there was no statistically significant difference between the guided discovery condition (21.22, 488.00) and the worked example condition (25.78, 593.00), U (23, 23) = 212.00, $p > .05$.

There was no statistically significant difference on the delayed post-assessment of the constructed mental model between the guided discovery condition (24.87, 572.00) and the worked example condition (22.13, 509.00), *U* (23, 23) = 233.00, *p* > .05.

Since there were no statistically significant differences between the two conditions at t2, the difference scores between t2 (delayed post-tests) and t1 (post-tests) were calculated to test if there was a statistically significant difference in decline between the two conditions. These difference scores for both conditions are shown in Table 6, by using the Mann-Whitney test.

Table 6. *Difference scores (t2-t1) comparing both conditions.*

	Condition					
	Guided discovery (n = 23)		Worked example (n = 23)			
Variable	Mean rank	Sum of ranks	Mean rank	Sum of ranks	Mann-Whitney U	*p*
Difference score total test	27.37	629.50	19.63	451.50	175.50	.05
Difference score retention test	28.57	657.00	18.43	424.00	148.00	.01
Difference score application test	24.74	569.00	22.26	512.00	236.00	.53
Difference score assessment of constructed mental model	24.61	566.00	22.39	515.00	239.00	.58

The guided discovery condition (27.37, 629.50) declined statistically significantly less on the total test than the worked example condition (19.63, 415.50), *U* (23, 23) = 175.50, *p* = .05.

On the retention test, the guided discovery condition (28.57, 657.00) declined statistically significantly less than the worked example condition (18.43, 424.00), *U* (23, 23) = 148.00, *p* = .01.

On the application test, there was no statistically significant difference in decline between the guided discovery condition (24.74, 569.00) and the worked example condition (22.26, 512.00), *U* (23, 23) = 236.00, *p* > .05.

There was no statistically significant difference in decline on the assessment of the constructed mental model between the guided discovery condition (24.61, 566.00) and the worked example condition (22.39, 515.00), *U* (23, 23) = 239.00, *p* > .05.

8.5. Is there a Difference in Experienced Cognitive Load During the Acquisition Phase, and Immediately After the Acquisition Phase, between the Guided Discovery and Worked Example Condition?

During the acquisition phase, the guided discovery condition (32.94, 790.50) experienced statistically significantly more cognitive load than the worked example condition (16.06, 385.50), U (24, 24) = 85.50, p = .00, see Table 5.

Immediately after the acquisition phase, the guided discovery condition (30.38, 729.00) experienced statistically significantly more cognitive load than the worked example condition (18.62, 447.00), U (24, 24) = 147.00, p = .00, see Table 5.

8.6. Is there a Difference in Motivation, Immediately After the Acquisition Phase, between the Guided Discovery and Worked Example Condition?

There was no statistically significant difference in motivation between the guided discovery condition (26.25, 630.00) and the worked example condition (22.75, 546.00), U (24, 24) = 246.00, p > .05, see Table 5.

8.7. Can the Learning Outcomes be Explained by Averaged Experienced Cognitive Load?

The averaged experienced cognitive load was calculated by averaging the experienced cognitive load during the acquisition phase with the experienced cognitive load immediately after the acquisition phase. The correlations between the averaged experienced cognitive load and the (delayed) post-tests for both conditions are shown in Table 7, by using the Spearman’s rankcorrelation.

Table 7. *Correlations between averaged experienced cognitive load and (delayed) post-tests for both conditions.*

	Condition			
	Guided discovery		Worked example	
Variable	Averaged experienced cognitive load	*p*	Averaged experienced cognitive load	*p*
Total post-test	-.53	.01	.08	.71
Retention post-test	-.45	.03	.12	.58
Application post-test	-.58	.00	.11	.61
Post-assessment of constructed mental model	.20	.36	.00	.99
Total delayed post-test	-.52	.01	.24	.26
Retention delayed post-test	-.53	.01	.16	.46
Application delayed post-test	-.46	.03	.35	.10
Delayed post-assessment of constructed mental model	-.09	.69	.14	.54

8.7.1. Correlations between averaged experienced cognitive load and the post-tests.

In the guided discovery condition there was a statistically significantly negative correlation between the averaged experienced cognitive load and the total post-test (*rho* = -.53, *p* = .01), the retention post-test (*rho* = -.45, *p* = .03) and the application post-test (*rho* = -.58, *p* = .00). There was no statistically significant correlation between the averaged experienced cognitive load and the post-assessment of the constructed mental model (*rho* = .20, *p* > .05).

In the worked example condition there was no statistically significant correlation between the averaged experienced cognitive load and the total post-test (*rho* = .08, *p* > .05), the retention post-test (*rho* = .12, *p* > .05), the application post-test (*rho* = .11, *p* > .05) and the post-assessment of the constructed mental model (*rho* = .00, *p* > .05).

8.7.2. Correlations between averaged experienced cognitive load and the delayed post-tests.

In the guided discovery condition there was a statistically significantly negative correlation between the averaged experienced cognitive load and the total delayed post-test (*rho* = -.52, *p* = .01), the retention delayed post-test (*rho* = -.53, *p* = .01) and the application delayed post-test (*rho* = -.46, *p* = .03). There was no statistically significant correlation between the averaged experienced cognitive load and the delayed post-assessment of the constructed mental model (*rho* = -.09, *p* > .05).

In the worked example condition there was no statistically significant correlation between the averaged experienced cognitive load and the total delayed post-test (*rho* = .24, $p > .05$), the retention delayed post-test (*rho* = .16, $p > .05$) and the delayed post-assessment of the constructed mental model (*rho* = .14, $p > .05$). The positive correlation between averaged experienced cognitive load and the application delayed post-test indicated a trend, *rho* = .35, $p = .10$.

8.8. Can the Learning Outcomes be Explained by Motivation?

Table 8. shows the correlations between motivation and the (delayed) post-tests for both conditions, by using Spearman's rankcorrelation.

Table 8. *Correlations between motivation and (delayed) post-tests for both conditions.*

	Condition			
	Guided discovery		Worked example	
Variable	Motivation	*p*	Motivation	*p*
Total post-test	.75	.00	-.01	.95
Retention post-test	.61	.00	-.10	.65
Application post-test	.70	.00	.02	.93
Post-assessment of constructed mental model	.14	.50	.08	.70
Total delayed post-test	.66	.00	-.14	.26
Retention delayed post-test	.55	.01	-.08	.71
Application delayed post-test	.67	.01	-.19	.40
Delayed post-assessment of constructed mental model	.16	.48	.26	.22

8.8.1. Correlations between motivation and the post-tests.

In the guided discovery condition there was a statistically significantly positive correlation between motivation and the total post-test (*rho* = .75, $p = .00$), the retention post-test (*rho* = .61, $p = .00$) and the application post-test (*rho* = .70, $p = .00$). There was no statistically significant correlation between motivation and the post-assessment of the constructed mental model (*rho* = .14, $p > .05$).

In the worked example condition there was no statistically significant correlation between motivation and the total post-test (*rho* = -.01, $p > .05$), the retention post-test (*rho* = -.10, $p > .05$), the application post-test (*rho* = .02, $p > .05$) and the post-assessment of the constructed mental model (*rho* = .08, $p > .05$).

8.8.2. Correlations between motivation and the delayed post-tests.

In the guided discovery condition there was a statistically significantly positive correlation between motivation and the total delayed post-test (*rho* = .66, *p* = .00), the retention delayed post-test (*rho* = .55, *p* = .01) and the application delayed post-test (*rho* = .67, *p* = .01). There was no statistically significant correlation between motivation and the delayed post-assessment of the constructed mental model (*rho* = .16, *p* > .05).

In the worked example condition there was no statistically significant correlation between motivation and the total delayed post-test (*rho* = -.14, *p* > .05), the retention delayed post-test (*rho* = -.08, *p* > .05), the application delayed post-test (*rho* = -.19, *p* > .05) and the delayed post-assessment of the constructed mental model (*rho* = .26, *p* > .05).

8.9. Correlation between Averaged Experienced Cognitive Load and Motivation

Since there was a statistically significant correlation between averaged experienced cognitive load and the (delayed) total, retention and application post-tests and between motivation and the (delayed) total, retention and application post-tests in the guided discovery condition, it was further investigated if there was a correlation between averaged experienced cognitive load and motivation.

In the guided discovery condition there was a statistically significantly negative correlation between averaged experienced cognitive load and motivation (*rho* = -.46, *p* = .02).

In the worked example condition there was a statistically significantly negative correlation between averaged experienced cognitive load and motivation (*rho* = -.46, *p* = .02).

9. Discussion

The aim of this study was to explore what the best way to learn the triage procedure is by comparing a serious game based on guided discovery learning with a PowerPointpresentation based on worked examples.

9.1. Learning Results

Both the serious game based on guided discovery learning and the PowerPointpresentation based on worked examples have the opportunity to teach students the triage procedure, since in both conditions there were statistically significant differences between all the (delayed) post-tests and the pre-tests. In other words, students can learn the triage procedure from a serious game based on guided discovery learning and also from a PowerPointpresentation based on worked examples.

9.2. Learning Outcomes in the Short Term

In the short term, immediately after the acquisition phase, there is a disadvantage of the guided discovery condition, since they scored statistically significantly lower on the total, retention and application post-test and experienced statistically significantly more cognitive load compared with the worked example condition. In other words, participants of the guided discovery condition reached lower learning outcomes in the short term while investigating more mental effort.

The learning outcomes in the guided discovery condition can possibly be explained by cognitive load theory, since it is assumed that the discovery process itself causes a high extraneous cognitive load, whereby there is not enough cognitive capacity available for germane cognitive load, or in other words learning, resulting in low learning outcomes. In addition, the learning outcomes in the worked example condition can possibly be explained by cognitive load theory, since it is assumed that worked examples cause a low extraneous cognitive load resulting in more cognitive capacity available to process the worked examples in order to construct schemas thereof. This results in high learning outcomes. Furthermore, these results are similar to previous research to the worked example effect that find an advantage of worked examples over conventional problem solving in the short term for different domains (Carroll, 1994; Cooper & Sweller, 1987; Kalyuga et al., 2001; Kalyuga et al., 2001; Lee et al., 2004; Paas, 1992; Paas & Van Merriënboer, 1994; Schwonke et al., 2009; Sweller & Cooper, 1985). However, it is questionable if the learning outcomes in the worked example condition can be explained by cognitive load theory, since there were no statistically significant correlations between averaged experienced cognitive load and the post-tests in this condition. Furthermore, the results of the present study are in contrast with theories about (un)guided discovery learning, since advocates of these theories argue that people learn the best when they discover and construct information by themselves.

The fact that there was no statistically significant difference between the two conditions on the post-assessment of the constructed mental model can possibly be explained by the duration of the acquisition phase. The maximum duration of the serious game was 17 minutes and the maximum duration of the PowerPointpresentation was 14 minutes, but according to Kahler (2001) a mental model is developed through experience and continued rehearsal of routine activities. Therefore it is possible that the acquisition phase was not long enough to reach a statistically significant difference between the two conditions on the post-assessment of the constructed mental model. For further research it is recommended to lengthen the acquisition phase to investigate if a statistically significant difference between the two conditions can be found.

Another possible explanation is that only a subset of the triage procedure was used to assess the constructed mental model, since using the whole triage procedure resulted in an unreliable assessment of the constructed mental model. It is the question if a statistically significant difference between the two conditions will be found if the whole triage procedure was used to assess the constructed mental model.

9.3. Learning Outcomes in the Long Term

For learning and instruction the learning outcomes in the long term are more interesting than the learning outcomes in the short term. In the long term, a week after the acquisition phase, there were no statistically significant differences between the two conditions. These results suggest that guided discovery learning and learning from worked examples give similar learning outcomes in the long term. However, investigating the differences between the delayed post-tests in the long term and the post-tests in the short term gives a somewhat different picture. Participants in the guided discovery condition declined statistically significantly less than participants in the worked example condition on the total and retention test. These results may suggest that guided discovery learning is more effective in the long term. These results are similar with the assumptions of (un)guided discovery learning, namely that having students discover and construct their own solutions to a problem or task leads to the most effective learning experience; that a greater amount of self-generated processing results in superior memory and performance of the target skill than a more passive encoding method; that students in (un)guided learning environments acquire more elaborated or robust knowledge, which is better understood, better organized and more amenable to future use, than students who merely read worked-out solutions; and that (un)guided discovery learning facilitates long-term retention (Clark et al., in press; Driscoll, 2005; Kebtrichi & Hirumi, 2008; Kirschner et al., 2006; Reiser et al., 1998).

Furthermore, it is possible that participants of the guided discovery condition remain more active in schema construction after the acquisition phase compared with participants in the worked example condition, resulting in more elaborated schemas which are advantageous in the long term.

In further research, it is interesting to investigate the learning outcomes in the long, long term by using a second delayed post-test to test if the participants in the guided discovery condition still declined less than participants in the worked example condition. Or to test if there is a statistically significant difference between the two conditions on the second delayed post-tests.

9.4. Experienced Cognitive Load

During and immediately after the acquisition phase, participants of the guided discovery condition experienced statistically significantly more cognitive load than participants of the worked example condition. A possible explanation is that participants of the guided discovery condition have to investigate cognitive resources in the discovery process itself, such as searching and finding out what the regularities and relationships are, and in learning the triage procedure. This investigation can be seen as germane cognitive load, since it contributes to learning (Sweller et al., 1998). Furthermore, they have to investigate cognitive resources in for example navigation in the game and searching of the victims, which are complex parts of playing serious games (Wouters et al., 2008). This investigation can be seen as extraneous cognitive load, since it is not intrinsic to the learning material and does not contribute to schema construction (Sweller et al., 1998). However, it is interesting to note that these complex parts are often intrinsic of serious games, but can be altered by game designers. In contrast, participants of the worked example condition only have to devote their cognitive resources to the studying of the appropriate steps necessary to reach the solution in order to construct and automate schemas, which caused statistically significantly less cognitive load.

Since there was a statistically significantly negative correlation between the averaged experienced cognitive load and the learning outcomes on the total, retention and application (delayed) post-tests in the guided discovery condition, it is interesting to further investigate which part of the cognitive load is caused by the learning process (germane cognitive load) and which part is caused by the design of the game (extraneous cognitive load), since a positive correlation between germane cognitive load and learning outcomes is desirable. A possible way to do this is using a think-aloud procedure where participants are asked to note experienced problems with either the learning process, the design of the game or both. However, it should be noted that in this case the participant also has to devote cognitive resources in thinking aloud and that this can interfere with the learning process and influence the learning outcomes. This suggestion for further research has practical implications. When the cognitive load is primarily caused by the learning process, it is the task of instructional designers to optimize the germane cognitive load, by for example offering more guidance such as cognitive load theorists suggest. However, when the cognitive load is primarily caused by the design of the game, it is the task of game designers to use design guidelines which reduce the extraneous cognitive load, thereby freeing working memory capacity for germane cognitive load with the intention to improve learning. Research in this domain is done by, among others, Van der Spek and colleagues (2009). A possible way to increase the germane cognitive load is using self-explanation prompts which require students to provide an explanation for each step they choose (Schwonke et al., 2009). It is important to note that increasing germane cognitive load is only possible when either intrinsic or extraneous cognitive load is low or when a combination of both applies (Sweller et al., 1998).

In addition, it is desirable to reduce the total cognitive load, since there was a statistically significantly negative correlation between averaged experienced cognitive load and motivation and a statistically significantly positive correlation between motivation and learning outcomes on the total, retention and application (delayed) post-tests in the guided discovery condition.

When it is clear what causes the cognitive load it is possible to improve the serious game Code Red: Triage and to replicate the present study to investigate if different results are found.

In contrast to the guided discovery condition, averaged experienced cognitive load had no statistically significant correlation with the learning outcomes on the (delayed) post-tests in the worked example condition. A possible explanation is that worked examples resemble more traditional instruction with which the participants are more familiar compared with serious gaming.

9.5. Motivation

Despite one of the main reasons to use serious games in education is the assumed motivation for students, no statistically significant difference was found between the conditions on the dependent variable motivation. This corresponds with the review studies of Wouters and colleagues (2009) and Hays (2005), who find no convincing evidence for the assumed motivation of serious games. However, this result is in contrast with the review study of Randel and colleagues (1992) who find that students in 12 out of 14 studies reported more interest in simulations/games activities than in more conventional classroom instruction.

Possible explanations of the finding in the present study are that participants in the guided discovery condition experienced a high cognitive load and some participants were frustrated that they could not find all the victims and/or had some trouble with the navigation of the game. Furthermore, some participants commented that they do not like games at all. Therefore it is recommended for further research to ask participants about their general motivation to play games and potentially use this finding as a covariate.

In the guided discovery condition, motivation had a statistically significantly positive correlation with the learning outcomes on the total, retention and application (delayed) post-tests. These results correspond with the assumption that increased motivation will lead to more learning and performance (Clark et al., in press). A practical implication from these results is that it is important for game designers to design serious games which are motivating for students in order to increase the learning outcomes. However, game designers should be aware of the fact that some students do not like games at all; games are not motivating for them and it is possible that they learn better from another medium, which is more motivating for them.

In contrast to the guided discovery condition, motivation had no statistically significant correlation with the learning outcomes on the (delayed) post-tests in the worked example condition. A possible explanation is that worked examples resemble more traditional instruction with which the participants are more familiar compared with serious gaming.

9.6. Limitations of the Present Study

It is the question to what extent the results of the present study are generalizable, since this study has some limitations. First, due to time constraints and the willingness of participants to participate, the used sample is rather small and not normally distributed, which makes it questionable to what extent the results are generalizable to the whole population of students. Furthermore, it is the question to what extent the results are generalizable to other parts of the population, such as adults who are not part of the new generation of students, such as described by Oblinger (2004). In addition, the participants of this study are novices. However, it is interesting to investigate if the same results are found with more experienced participants.

Second, in this study the only learning objective of the instruction was learning the triage procedure, which is part of the medical domain. This makes it difficult to say if the results are generalizable to other domains. Therefore it is interesting to further investigate if the same results are found by comparing a serious game based on guided discovery learning with worked examples, when the participants have to learn another procedure in another domain. In addition, it is interesting to further investigate if serious games based on guided discovery learning have the potential to increase learning of facts, concepts and principles, which are also part of the cognitive domain (Merrill, 1994) when compared with worked examples. It should be noted, that it is also interesting to investigate if serious games based on guided discovery learning have the potential to increase learning in the affective and psychomotor domains. However, a comparison with worked examples is not so obvious then.

Third, in this study a comparison is made between guided discovery learning and worked examples, but there are a lot of different operationalizations and designs of these concepts possible. Therefore it is questionable if another researcher will find the same results when using another operationalization and design of the learning environments.

Fourth, the paper-and-pencil tests used in this study resemble more the worked example condition than the guided discovery condition. It is possible that this has influenced the results, since an alignment between instruction and assessment is important (Pellegrino, 2004). Therefore it is interesting to investigate the learning outcomes when the tests are designed as a serious game and thus resemble the guided discovery condition more.

Fifth, this study was set up as an experiment in a laboratory setting. Therefore it is questionable to what extent the results are generalizable to authentic classroom situations where serious games and/or worked examples are often part of the whole instructional process. Therefore it is interesting to replicate this study in a more authentic classroom situation with an intended target of students, such as medical first responders.

10. Conclusion

An increasing number of researchers and educators is considering the use of serious games in education of the new generation of students. From the results of this study it is concluded that, in the case of learning the triage procedure with novice students, a serious game based on guided discovery learning is disadvantageous for learning outcomes in the short term compared with worked examples. However, in the long term participants who played the serious game declined less than participants in the worked example condition. This may suggest that serious games based on guided discovery learning are promising for long term learning, which is the intention of learning and instruction.

Since both experienced cognitive load and motivation are correlated with the short and long term learning outcomes when playing the game, further research is necessary to investigate how cognitive load and motivation can be optimized when playing games in order to increase learning.

Furthermore, interpretation and potentially generalization of the results of this study should take into account the mentioned limitations of this study.

Total amount of words: 12.992

References

Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology, 3,* 360-367.

Clark, R., Yates, K., Early, S., & Moulton, K. (in press). 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, Volume I: Instructional design and training delivery.* Somerset, NJ: Wiley.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 4,* 347-362.

De Waard, D. (1996). *The measurement of drivers' mental workload.* PhD thesis, University of Groningen. Haren: The Netherlands: University of Groningen, Traffic Research Centre.

Driscoll, M. P. (2005). *Psychology of learning for instruction.* Boston: Pearson Education.

Gagné, R. M. (1971). The learning of principles. In M. D. Merrill (Ed.), *Instructional design: readings* (pp. 303-315). New Jersey: Prentice-Hall.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33,* 441-467.

Gray, P. (2007). *Psychology, Fifth edition.* New York: Worth Publishers.

Halff, H. M. (2005). Adventure games for science education: Generative methods in exploratory environments. *Proceedings of AIED05 Workshop,* 1-12.

Hays, R. T. (2005). *The effectiveness of instructional games: A literature review and discussion* (Technical Report, 2005-004). Orlando, FL: Naval Air Warfare Center Training Systems Divisions.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42,* 99-107.

Hustinx, P., Meeuwis, D., & Hermans, R. *Geneeskundig management bij grootschalige incidenten* [Medical management by major incidents]. Utrecht: De Tijdstroom.

Kahler, S. E. (2001). *A comparison of knowledge acquisition methods for the elicitation of procedural mental models.* Doctoral dissertation. North Carolina: North Carolina State University.

Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology, 1,* 5-23.

Kalyuga, S., Chandler, P., Sweller, J., & Tuovinen, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 3,* 579-588.

Kebtritchi, M., & Hirumi, A. (2008). Examining the pedagogical foundations of modern educational computer games. *Computers & Education, 51,* 1729-1743.

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *Internet & Higher Education, 8,* 13-24.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41,* 75-86.

Lau, W. W. F., & Yuen, A. H. K. (2010). Promoting conceptual change of learning sorting algorithm through the diagnosis of mental models: The effects of gender and learning styles. *Computers and Education, 54,* 275-288.

Lee, K. M., Nicoll, G., & Brooks, D. W. (2004). A comparison of inquiry and worked example web-based instruction using physlets. *Journal of Science Education and Technology, 1,* 81-88.

Mayer, R. E. (2004). Should there be a three-strikes rule against discovery learning? The case for guided methods of instruction. *American Psychologist, 1,* 14-19.

Merrill, M. D. (1994). *Instructional design theory.* New Jersey: Educational Technology Publications.

Merrill, M. D. (1999). Instructional transaction theory (ITT): Instructional design based on knowledge objects. In C. M. Reigeluth (Ed.), *Instructional-design theories and models. A new paradigm of instructional theory. Volume II* (pp. 397 -425). New York and London: Routledge.

Oblinger, D. G. (2004). The next generation of educational engagement. *Journal of Interactive Media in Education, 8,* 1-12.

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 4,* 429-434.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 1,* 122-133.

Prensky, M. (2003). Digital game-based learning. *ACM Computers in Entertainment, 1,* 1-4.

Pellegrino, J. W. (2004). Complex learning environments: connecting learning theory, instructional design, and technology. In N. M. Seel, & S. Dijkstra (Eds.), *Curriculum, plans, and processes in instructional design* (pp. 25-48). New York: Lawrence Erlbaum Associates.

Randel, J. M., Morris, B. A., Wetzel, C. D., & Whitehil, B. V. (1992). The effectiveness of games for educational purposes: A review of recent research. *Simulation & Gaming, 23,* 261-275.

Reiser, B. J., Copen, W. A., Ranney, M., Hamid, A., & Kimberg, D. Y. (1998). *Cognitive and* *motivational consequences of tutoring and discovery learning* (ARI Research Note 98-12). Evanston, IL: School of Education and Social Policy.

Schank, R. C., Berman, T. R., & Macpherson, K. A. (1999). Learning by doing. In C. M. Reigeluth (Ed.), *Instructional-design theories and models. A new paradigm of instructional theory. Volume II* (pp. 161-181). New York and London: Routledge.

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior, 25,* 258-266.

Schmidt, H. G., Loyens, S. M. M., Van Gog, T., & Paas, F. (2007). Problem-based learning *is* compatible with human cognitive architecture: commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42,* 91-97.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (19-30). New York: Cambridge University Press.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 1,* 59-89.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 3,* 251-296.

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 2,* 334-341.

Van Merriënboer, J. G., & Kirschner, P. A. (2007). *Ten steps to complex learning. A systematic approach to four-component instructional design.* Mahwah: Lawrence Erlbaum Associates.

Van der Spek, E. D., Van Oostendorp, H., Wouters, P. & Aarnoudse, L. (2010). Attentional cueing in serious games. In *IEEE 2nd International Conference on Games and Virtual Worlds for Serious Applications* (pp. 199-125).

Van der Spek, E. D., Wouters, P., & Van Oostendorp, H. (2009). Code Red: Triage. Or, COgnition-Based DEsign Rules Enhancing Decisionmaking TRaining in a Game Environment. In *IEEE Conference in Games and Virtual Worlds for Serious Applications* (pp. 166-169).

Wouters, P., Van der Spek, E. D., & Van Oostendorp, H. (2008). Serious games for crisis management: What can we learn from research on animations? In A. Maes & S. Ainsworth (Eds.), *Exploiting the Opportunities: Learning with Textual, Graphical and Multimodal Representations* (pp. 162-165). Tilburg: Earli.

Wouters, P., Van der Spek, E. D., & Van Oostendorp, H. (2009). Current practices in serious game research: A review from a learning outcomes perspective. In T. Connolly, M. Stansfield, & L. Boyle (Eds.), *Games-based learning advancements for multi-sensory human computer interfaces: techniques and effective practices* (pp. 232-250). Hersey: IGI Global.

Appendix A. Explanation of the guided discovery condition.

De triage van slachtoffers

Wanneer er een groot ongeluk is gebeurd waarbij veel slachtoffers zijn gevallen, moet de eerste ambulancebroeder ter plaatse bepalen welke slachtoffers het dringendst medische hulp nodig hebben. Tijdens deze categorisatie, de zogeheten primaire triage, moeten de slachtoffers ondergebracht worden in één van de vier triagecategorieën. De categorieën zijn **T3**: niet urgent; **T2**: redelijk urgent; **T1**: hoogst urgent; en **Dood**. De primaire triage vindt plaats op de plek waar de slachtoffers gevonden worden. Het onderbrengen van een slachtoffer in een van de categorieën dient zo snel mogelijk te gebeuren.

Uitleg over de game Code Red: Triage

In de game Code Red: Triage ga je de procedure van de primaire triage leren. De procedure is voor alle slachtoffers hetzelfde, en begint bij elk slachtoffer bij dezelfde stap. Het is aan jou de taak om te achterhalen hoe deze procedure gaat.

Het doel van de game is om erachter te komen wat de procedure van de primaire triage is en deze vervolgens te oefenen. Het gaat erom dat je de procedure zo snel en zo correct mogelijk uitvoert.

De game begint in een stationshal. Daar krijg je informatie over wat er gebeurd is en waar je de slachtoffers kunt vinden. Om de slachtoffers te vinden moet je eerst door het gangenstelsel van het metrostation lopen. In de game krijg je hierover aanwijzingen. Dit deel van de game is bedoeld om te oefenen hoe je je in de game kunt voortbewegen. Je kunt de volgende toetsen gebruiken:

Toets	
W	**naar voren gaan**
A	**naar links gaan**
D	**naar rechts gaan**
S	**achteruit gaan**

Met de muis kun je de richting bepalen. Tijdens het lopen door het gangenstelsel wordt twee keer een nieuw ‘hoofdstuk’ van de game geladen. Dat duurt enkele seconden.

Je hebt 17 minuten de tijd om alle slachtoffers in een triagecategorie in te delen. Zodra je het eerste slachtoffer bereikt, begint het aftellen. **Wanneer je bij een slachtoffer aankomt, kun je een triage uitvoeren door op de *e-toets* te drukken.** Dan verschijnt een interface met links en rechts een aantal buttons (zie figuur 1). In het begin zullen dit slechts een paar buttons zijn, naarmate je verder komt, komen er steeds meer buttons bij.

Figuur 1. Interface met links en rechts een aantal buttons.

Standaard verschijnt introductie informatie: de leeftijd van het slachtoffer en eventueel de verwondingen. Met de buttons kun je extra informatie opvragen. Elke button geeft algemene uitleg over het onderwerp van de button en informatie over het slachtoffer met betrekking tot het onderwerp.
De button “Ademweg” kan bijvoorbeeld als algemene uitleg geven: “Je bepaalt of de ademweg vrij is en of het slachtoffer zelfstandig ademt. Als de ademweg niet vrij is, dient deze vrijgemaakt te worden.” De slachtoffer specifieke informatie zou kunnen zijn: “Het slachtoffer kan vrij ademen”. In de interface zijn deze twee onderdelen van elkaar gescheiden door een lege regel. **De algemene uitleg is altijd hetzelfde, maar geeft wel hints over hoe de triageprocedure verloopt.**

Omdat het meten van sommige waardes (zoals de polsslag) in werkelijkheid enige tijd kost, verschijnt deze informatie niet direct maar moet je even wachten.

Niet alle buttons zijn altijd nodig om slachtoffers in te delen in een triagecategorie. Welke buttons van toepassing zijn, hangt af van de triage procedure en de toestand van het slachtoffer. Tijd is een cruciale factor bij de primaire triage. Een button selecteren die niet nodig is, kost tijd. Daarom is het belangrijk om de juiste procedure te achterhalen en vervolgens alleen de buttons te gebruiken die nodig zijn.

Wanneer je genoeg informatie denkt te hebben, kun je met een van de vier buttons onder in het scherm de juiste triagecategorie selecteren. Het slachtoffer krijgt daarna de kleur van de gekozen triagecategorie (een T3 slachtoffer wordt dus groen), waarna het triagescherm afsluit en je weer door de virtuele wereld kunt bewegen.

Elke keer dat je een slachtoffer in een triagecategorie indeelt, krijg je kortstondig in beeld te zien wat je score voor het slachtoffer is. Verder krijg je informatie over of je:

a. de primaire triage binnen de voorgeschreven tijd hebt uitgevoerd;
b. de stappen van de procedure in de juiste volgorde hebt gedaan;
c. stappen bent vergeten;
d. overbodige stappen hebt gedaan

Je kunt 100 punten halen voor het indelen van het slachtoffer in de juiste triagecategorie, maar hiervan worden punten afgetrokken voor:

a. Stappen van de procedure in de verkeerde volgorde doen;
b. belangrijke stappen vergeten;
c. overbodige stappen doen; en/of
d. de voorgeschreven tijd overschrijden.

De totale score bovenin het scherm verandert. Daarna kun je naar het volgende slachtoffer.

Tijdens het spelen van de game verschijnt eenmalig een pop-up waarin je wordt gevraagd om een vraag op papier te beantwoorden.

Denk dus goed na over welke buttons je aanklikt voor een goede score! En bedenk: soms zijn andere buttons nodig, **maar de volgorde van stappen in de procedure blijft altijd hetzelfde, en de procedure begint altijd met dezelfde stap**!

Wanneer het duidelijk is hoe de game werkt, kun je de experimentbegeleidster vragen de game te starten. Indien je nog vragen hebt, kun je die ook aan haar voorleggen.

Appendix B. Explanation of the worked example condition.

De triage van slachtoffers

Wanneer er een groot ongeluk is gebeurd waarbij veel slachtoffers zijn gevallen, moet de eerste ambulancebroeder ter plaatse bepalen welke slachtoffers het dringendst medische hulp nodig hebben. Tijdens deze categorisatie, de zogeheten primaire triage, moeten de slachtoffers ondergebracht worden in één van de vier triagecategorieën. De categorieën zijn **T3**: niet urgent; **T2**: redelijk urgent; **T1**: hoogst urgent; en **Dood**. De primaire triage vindt plaats op de plek waar de slachtoffers gevonden worden. Het onderbrengen van een slachtoffer in een van de categorieën dient zo snel mogelijk te gebeuren.

Uitleg over de powerpoint Code Red: Triage

In de powerpoint Code Red: Triage ga je de procedure van de primaire triage leren. De procedure is voor alle slachtoffers hetzelfde, en begint bij elk slachtoffer bij dezelfde stap. Het is aan jou de taak om te achterhalen hoe deze procedure gaat.

Het doel van de powerpoint is om erachter te komen wat de procedure van de primaire triage is. Het gaat erom dat je dit zo snel en zo correct mogelijk doet.

Je hebt 14 minuten om alle slachtoffercasussen door te nemen. De powerpoint begint met informatie over wat er gebeurd is. Daarna volgen er een aantal slachtoffercasussen. Elke casus begint met introductie informatie: de leeftijd van het slachtoffer en eventueel de verwondingen. Daarna volgen de stappen die je volgens de primaire triage moet uitvoeren. Elke stap geeft algemene uitleg over het onderwerp van de stap en informatie over het slachtoffer met betrekking tot die stap. De stap “Ademweg” kan bijvoorbeeld als algemene uitleg geven: “Je bepaalt of de ademweg vrij is en of het slachtoffer zelfstandig ademt. Als de ademweg niet vrij is, dient deze vrijgemaakt te worden.” De slachtoffer specifieke informatie zou kunnen zijn: “Het slachtoffer kan vrij ademen”. De algemene uitleg is ***cursief en dikgedrukt***; de slachtoffer specifieke informatie staat op de regel na de algemene uitleg. **De algemene uitleg is altijd hetzelfde, maar geeft wel hints over hoe de triageprocedure verloopt.** Figuur 1 geeft een impressie van de powerpoint.

Figuur 1. Powerpoint.

Tijdens het lezen van de powerpoint verschijnt eenmalig een dia waarin je wordt gevraagd om een vraag op papier te beantwoorden.

Bedenk: soms zijn andere buttons nodig, afhankelijk van de toestand van het slachtoffer, **maar de volgorde van stappen in de procedure blijft altijd hetzelfde, en de procedure begint altijd met dezelfde stap!**

Wanneer het duidelijk is hoe de powerpoint werkt, kun je de experimentbegeleidster vragen de powerpoint te starten. Indien je nog vragen hebt, kun je die ook aan haar voorleggen.

Appendix C. Retention test.

Algemene gegevens

Naam:	

Onderdeel A.
Je krijgt tien als-dan zinnen. Vul elke zin aan met de actie die je volgens de primaire triage moet doen. Bij de vijf tekstuele vragen omcirkel je het juiste antwoord. Bij de vijf visuele vragen kruis je het juiste antwoord aan.

Als je nog vragen hebt over de toets, vraag deze dan aan de experimentleidster. Anders kan je beginnen. Je hebt vijf minuten de tijd.

Succes!

Vraag 1.
Als de capillaire refill tijd te hoog is, dan …

A. Controleer je de polsslag.
B. Controleer je de ademweg.
C. Deel je het slachtoffer in in triagecategorie T1.
D. Controleer je de ademhalingsfrequentie.
E. Deel je het slachtoffer in in triagecategorie T2.
F. Gebruik je de chin lift.

Vraag 2.

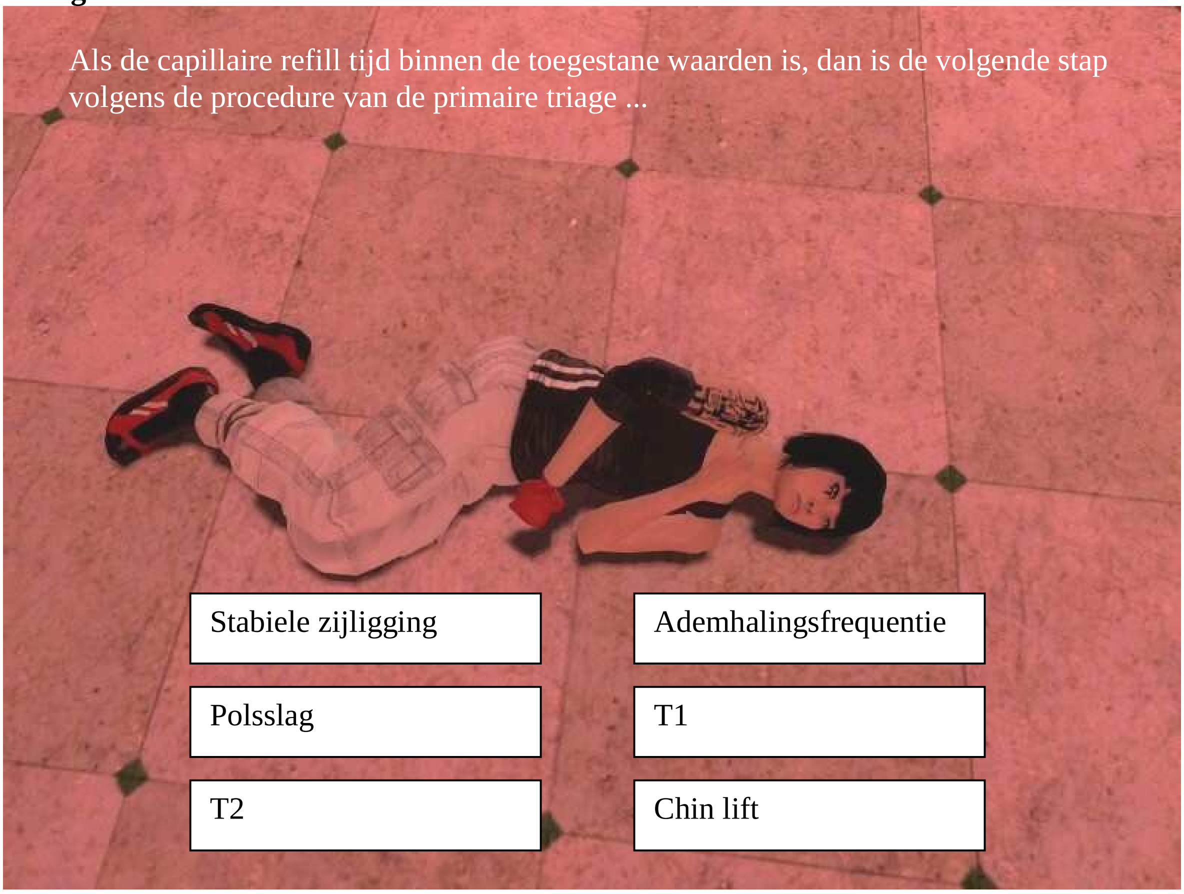
Vraag 3.
Als de ademhalingsfrequentie te laag of te hoog is, dan …

A. Deel je het slachtoffer in in triagecategorie T2.
B. Controleer je de mobiliteit.
C. Deel je het slachtoffer in in triagecategorie T1.
D. Controleer je de ademweg.
E. Gebruik je de chin lift.
F. Controleer je de capillaire refill tijd.

Vraag 4.

Vraag 5.
Als de polsslag te hoog is, dan …
- A. Gebruik je de chin lift.
- B. Gebruik je de stabiele zijligging.
- C. Deel je het slachtoffer in in triagecategorie T2.
- D. Controleer je de capillaire refill tijd.
- E. Controleer je de ademhalingsfrequentie.
- F. Deel je het slachtoffer in in triagecategorie T1.

Vraag 6.



Vraag 7.
Als de ademhalingsfrequentie binnen de toegestane waarden valt, dan …

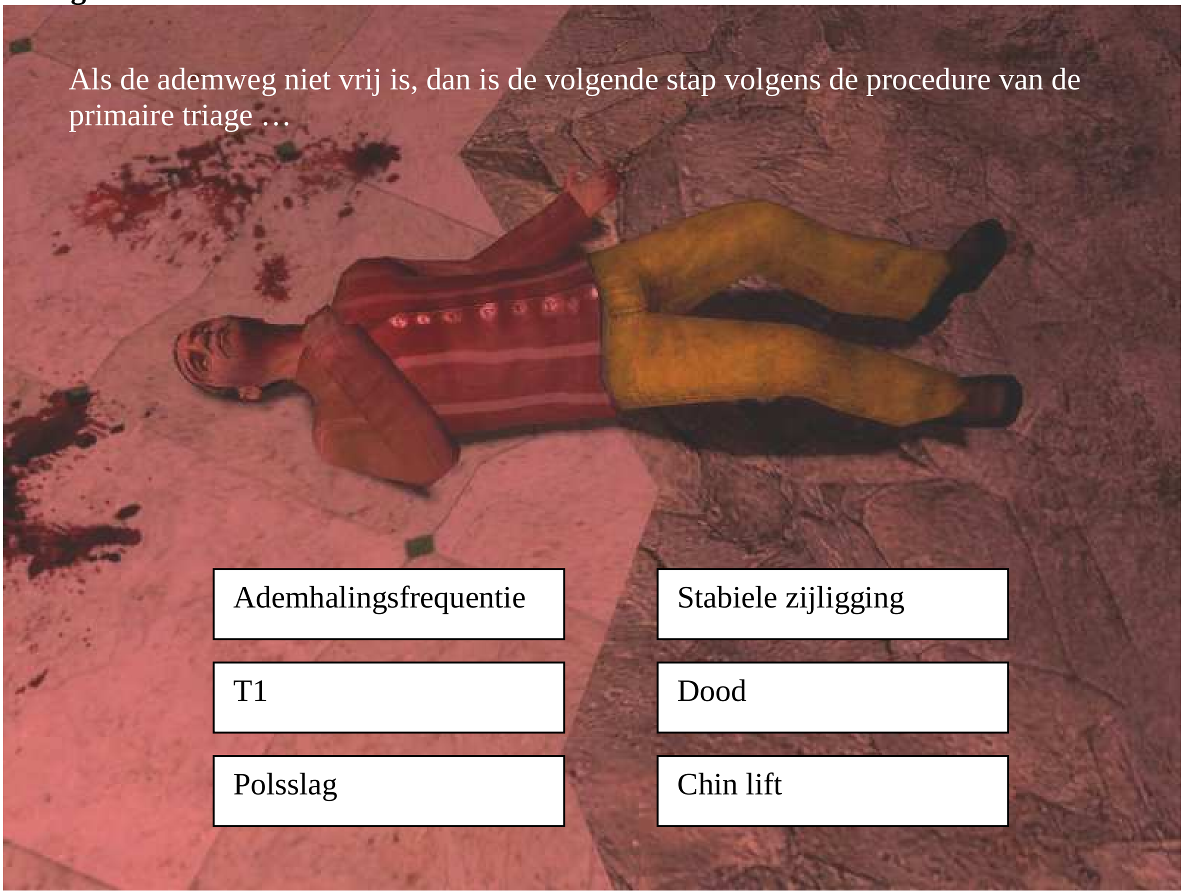
A. Controleer je de polsslag.
B. Deel je het slachtoffer in in triagecategorie T3.
C. Controleer je de capillaire refill tijd.
D. Controleer je de ademweg.
E. Gebruik je de chin lift.
F. Deel je het slachtoffer in in triagecategorie T2.

Vraag 8.

Vraag 9.
Als de ademweg vrij is, dan …

A. Controleer je de ademhalingsfrequentie.
B. Deel je het slachtoffer in in triagecategorie T3.
C. Controleer je de mobiliteit.
D. Controleer je de polsslag.
E. Gebruik je de jaw thrust.
F. Controleer je de capillaire refill tijd.

Vraag 10.



Einde onderdeel A.

Appendix D. Application test.

Algemene gegevens

Naam:	

Onderdeel B.
Je krijgt tien casussen. Kies bij elke casus de actie die je in die situatie volgens de procedure van de primaire triage moet doen. Bij de vijf tekstuele vragen omcirkel je het juiste antwoord. Bij de vijf visuele vragen kruis je het juiste antwoord aan.

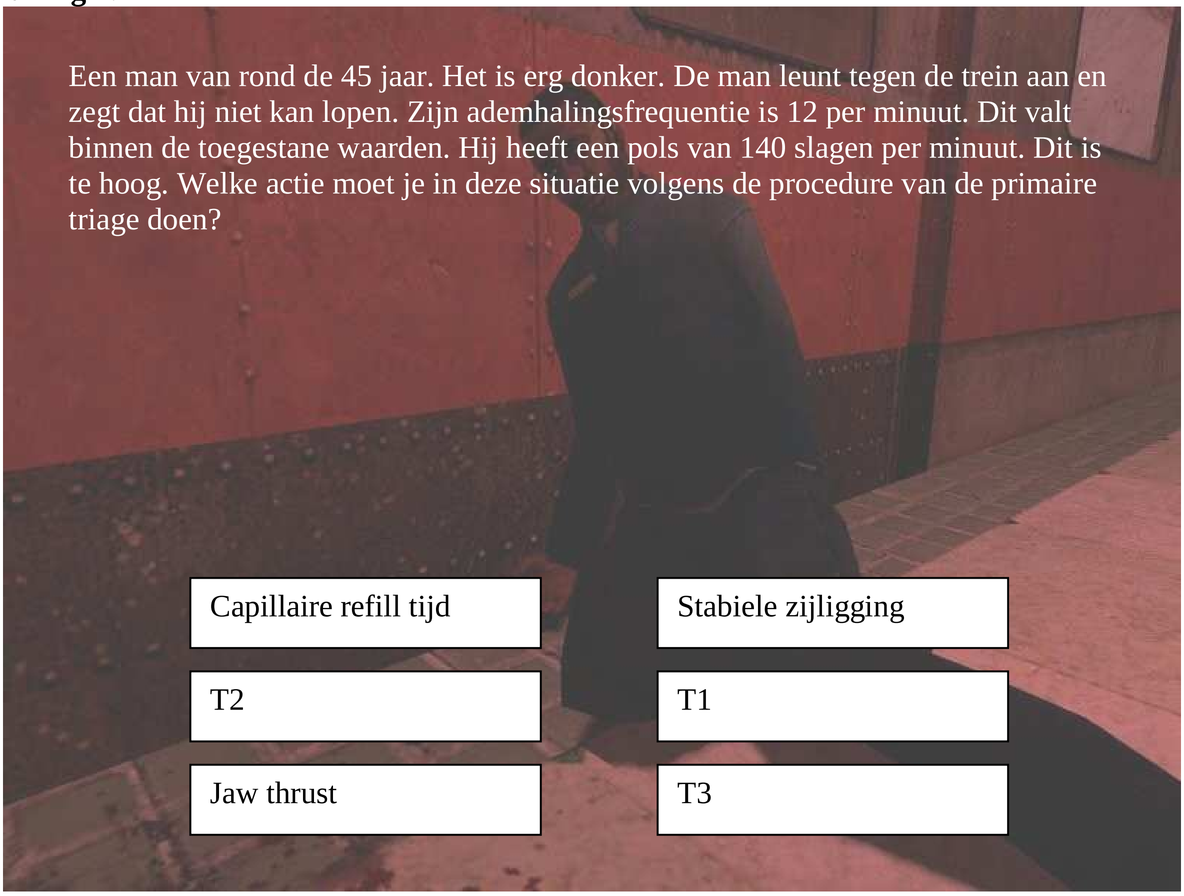
Als je nog vragen hebt over de toets, vraag deze dan aan de experimentleidster. Anders kan je beginnen. Je hebt tien minuten de tijd.

Succes!

Vraag 1.
Een man van rond de 30 jaar. Hij mist een been als gevolg van een explosie. Je hebt een geblokkeerde ademweg geconstateerd, maar nadat je deze hebt vrijgemaakt is de ademhaling niet op gang gekomen. Welke actie moet je in deze situatie volgens de procedure van de primaire triage doen?

A. Het slachtoffer indelen in triagecategorie dood.
B. De stabiele zijligging gebruiken.
C. De ademhalingsfrequentie controleren.
D. De capillaire refill tijd controleren.
E. Het slachtoffer indelen in triagecategorie T1.
F. De polsslag controleren.

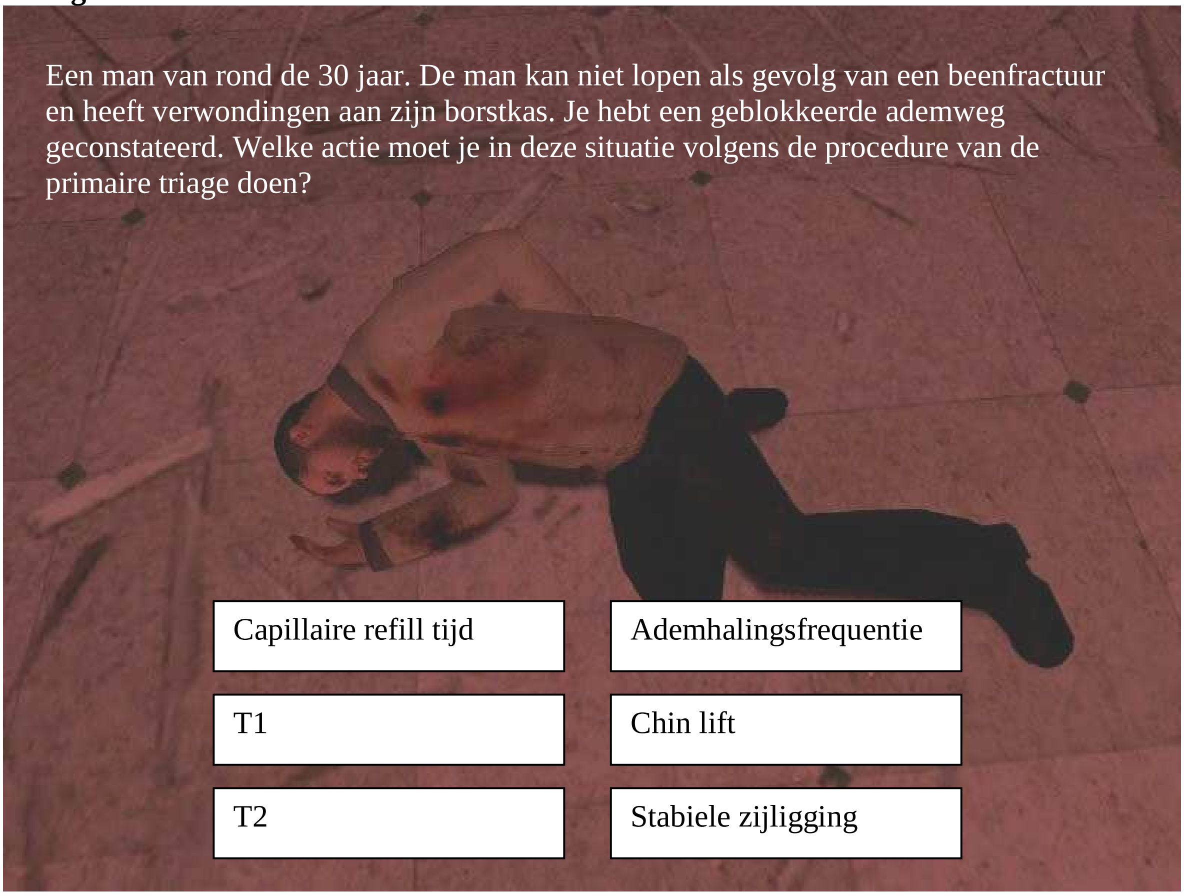
Vraag 2.



Vraag 3.
Een vrouw van rond de 20 jaar. Ze geeft aan dat ze door duizeligheid niet kan lopen. Ze kan vrij ademen en haar ademhalingsfrequentie is normaal. Welke actie moet je in deze situatie volgens de procedure van de primaire triage doen?

A. De stabiele zijligging gebruiken.
B. Het slachtoffer indelen in triagecategorie T3.
C. Het slachtoffer indelen in triagecategorie T2.
D. De polsslag controleren.
E. De capillaire refill tijd controleren.
F. De chin lift toepassen.

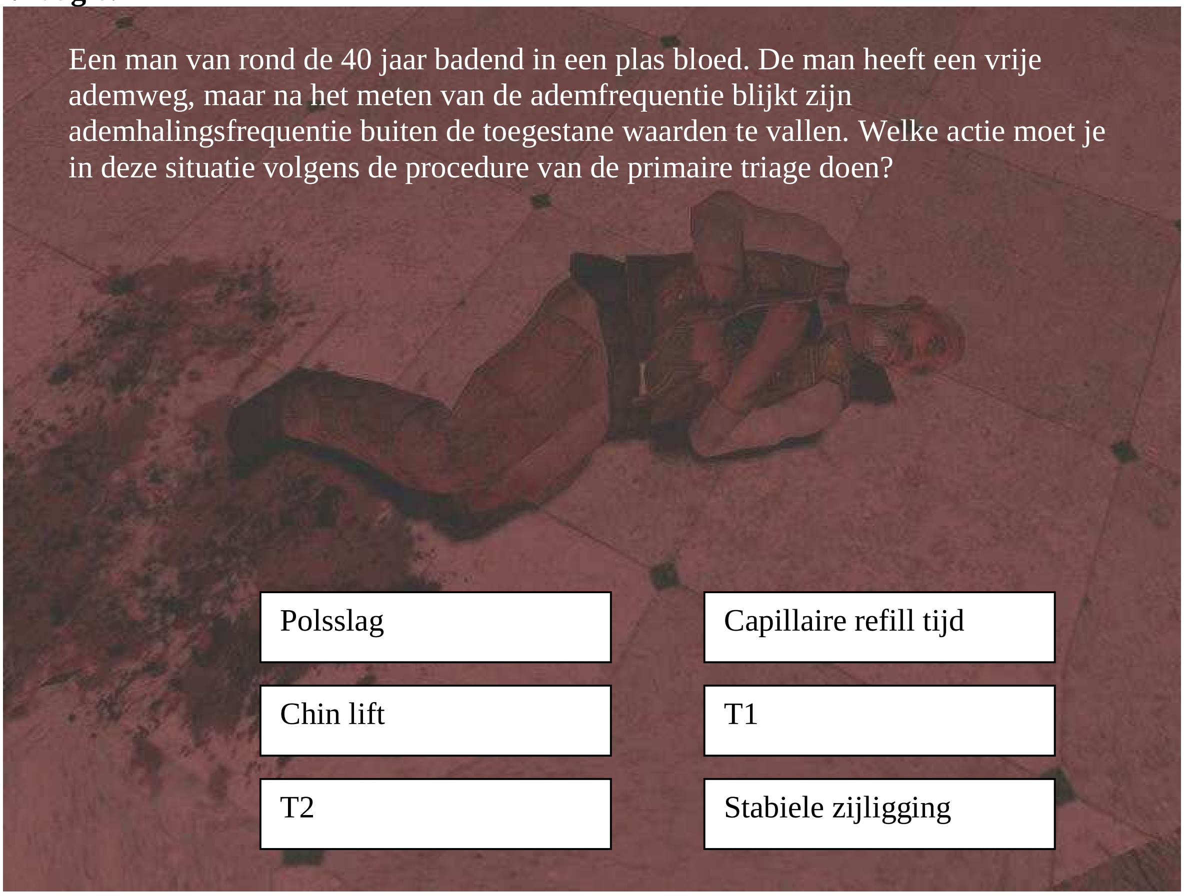
Vraag 4.



Vraag 5.
Een vrouw van rond de 20 jaar met verwondingen aan het bovenlichaam. Het slachtoffer kan niet lopen, maar heeft wel een vrije ademweg. Welke actie moet je in deze situatie volgens de procedure van de primaire triage doen?

A. De polsslag controleren.
B. De capillaire refill tijd controleren.
C. Het slachtoffer indelen in triagecategorie T1.
D. De ademhalingsfrequentie controleren.
E. De chin lift toepassen.
F. Het slachtoffer indelen in triagecategorie T2.

Vraag 6.

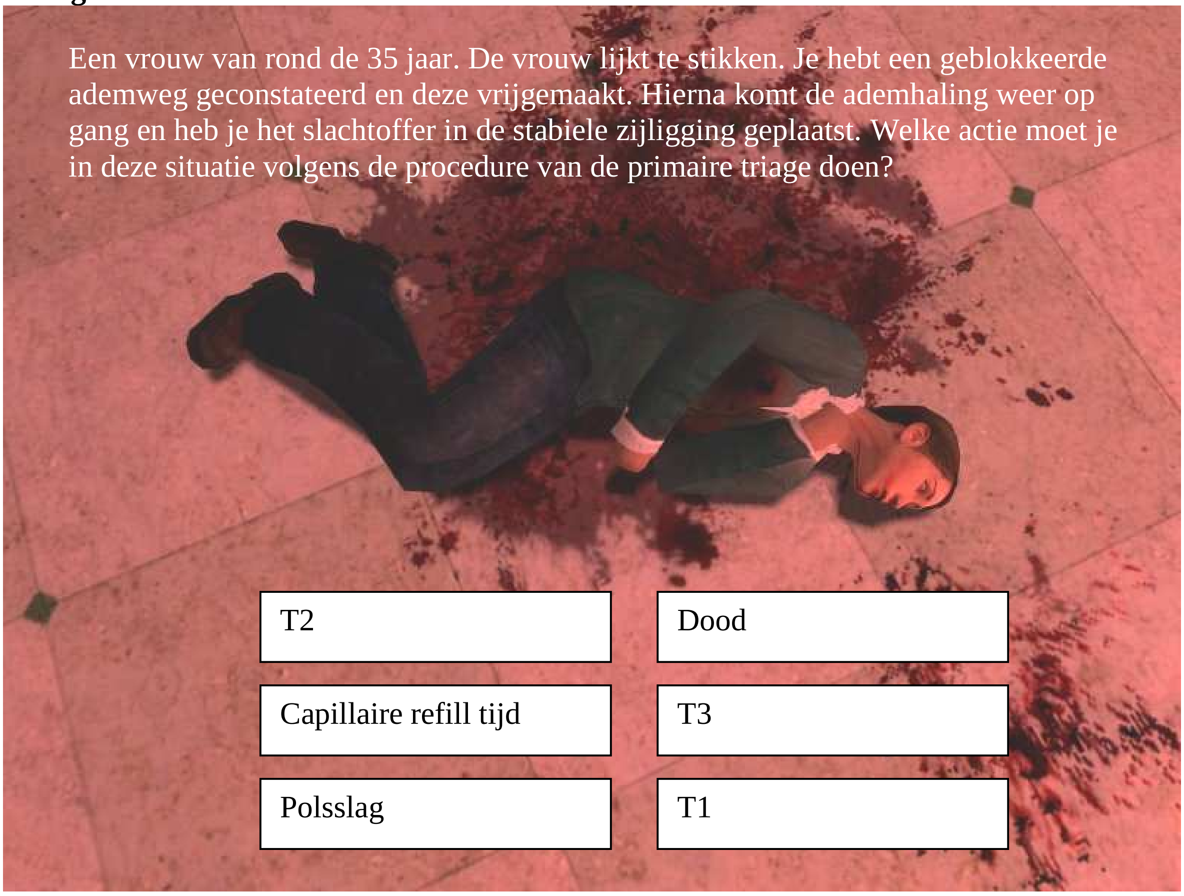


Vraag 7.

Een bewusteloze vrouw van rond de 50 jaar. De ademweg was geblokkeerd, maar je hebt deze zojuist vrijgemaakt en de ademhaling is weer op gang gekomen. Welke actie moet je in deze situatie volgens de procedure van de primaire triage doen?

A. De stabiele zijligging gebruiken.
B. Het slachtoffer indelen in triagecategorie T1.
C. De ademhalingsfrequentie controleren.
D. Het slachtoffer indelen in triagecategorie T2.
E. De capillaire refill tijd controleren.
F. De polsslag controleren.

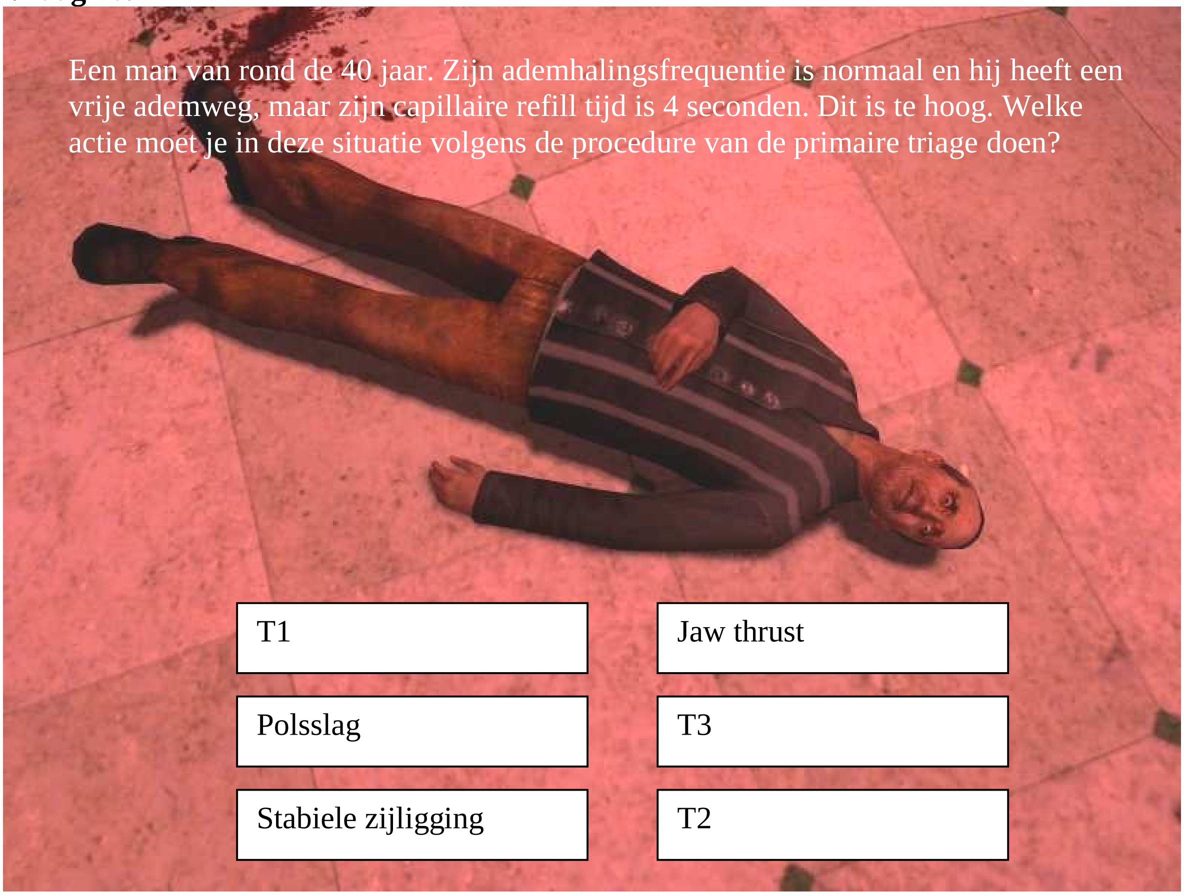
Vraag 8.



Vraag 9.
Een vrouw van rond de 45 jaar. Er is een stuk muur op haar been gevallen. Ze kan vrij ademen en zowel haar capillaire refill tijd als ademhalingsfrequentie zijn normaal. Welke actie moet je in deze situatie volgens de procedure van de primaire triage doen?

- A. De polsslag controleren.
- B. Het slachtoffer indelen in triagecategorie T2.
- C. De chin lift toepassen.
- D. Het slachtoffer indelen in triagecategorie T3.
- E. De stabiele zijligging toepassen.
- F. Het slachtoffer indelen in triagecategorie T1.

Vraag 10.



Einde onderdeel B.

Appendix E. Assessment of the constructed mental model.

Algemene gegevens

Naam:	
Geslacht:	
Leeftijd (in jaren):	
Huidig opleidingsniveau (HBO, WO):	
Studie:	
Telefoonnummer:	
E-mailadres:	

Uitleg
Zo dadelijk krijg je 28 paren aangeboden van concepten die te maken hebben met de beoordeling van een slachtoffer. We vragen je voor elk paar aan te geven in welke mate de concepten volgens jou aan elkaar gerelateerd zijn. Er zijn verschillende manieren waarop je over zulke relaties kan nadenken. Twee concepten kunnen bijvoorbeeld gemeenschappelijke kenmerken hebben of de twee concepten kunnen regelmatig gelijktijdig optreden. In dit experiment zijn we vooral geïnteresseerd in je eerste indruk bij het zien van de paren. Daarom vragen we je om je beoordeling vooral te baseren op deze indruk.

Elk paar met concepten wordt gepresenteerd met een schaal waarop je de mate van relatie kan weergeven. Deze schaal loopt van “1” tot en met “9”. Hoe hoger het getal, hoe groter de mate van relatie tussen twee concepten. Wanneer je vindt dat twee concepten heel sterk gerelateerd zijn, kan je een “8” of een “9” omcirkelen. Wanneer je vindt dat de concepten helemaal geen relatie hebben kan je een “1” of een “2” omcirkelen.

Je kan de mate van relatie aangeven door een getal te omcirkelen.

Hieronder zie je de lijst met concepten. Je kan dan een beeld krijgen van het soort concepten dat gepresenteerd wordt.

Lijst met concepten die beoordeeld moeten worden:

Chin lift	Mobiliteit
Slachtoffer classificeren	Stabiele zijligging
T3	T1
Lichtgewond	Geblokkeerde ademweg

Als je nog vragen hebt over de taak, vraag deze dan aan de experimentleidster. Anders kan je beginnen met het beoordelen van de concepten. Je hebt hiervoor vijf minuten de tijd.

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Slachtoffer classificeren
Stabiele zijligging

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

T3
Chin lift

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Geblokkeerde ademweg
Slachtoffer classificeren

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Chin lift
T1

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Geblokkeerde ademweg
T3

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

T3
Lichtgewond

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Slachtoffer classificeren
T3

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

T1
Slachtoffer classificeren

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Chin lift
Slachtoffer classificeren

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Stabiele zijligging
Mobiliteit

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

T1
Stabiele zijligging

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Chin lift
Geblokkeerde ademweg

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Chin lift
Stabiele zijligging

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Lichtgewond
Chin lift

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

T3
T1

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

T1
Geblokkeerde ademweg

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Geblokkeerde ademweg
Stabiele zijligging

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Stabiele zijligging
Lichtgewond

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Slachtoffer classificeren
Lichtgewond

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Mobiliteit
Chin lift

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Mobiliteit
T3

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Stabiele zijligging
T3

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Geblokkeerde ademweg
Mobiliteit

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

T1
Mobiliteit

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Lichtgewond
T1

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Slachtoffer classificeren
Mobiliteit

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Mobiliteit
Lichtgewond

Ongerelateerd	1	2	3	4	5	6	7	8	9	Gerelateerd

Geblokkeerde ademweg
Lichtgewond

Einde van dit onderdeel.