Computers & Education 70 (2014) 173–183

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

The design and effect of automated directions during scenario-based training

Marieke Peeters [a, b, *], Karel van den Bosch [b], John-Jules Ch. Meyer [a], Mark A. Neerincx [b, c]

[a] Intelligent Systems Group, Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands
[b] Training and Performance Innovations Group, Behavioral and Societal Sciences TNO, Soesterberg, The Netherlands
[c] Interactive Intelligence Group, Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Article history:
Received 16 October 2012
Received in revised form
22 July 2013
Accepted 23 July 2013

Keywords:
Intelligent agents
Scenario-based training
Director
Experiment
Didactical interventions
Learning value

ABSTRACT

During scenario-based training, the scenario is dynamically adapted in real time to control the storyline and increase its effectiveness. A team of experienced staff members is required to manage and perform the adaptations. They manipulate the storyline and the level of support during their role-play of important characters in the scenario. The costs of training could be reduced if the adaptation is automated by using intelligent agent technology to control the characters within a virtual training environment (a serious game). However, such a system also needs a didactical component to monitor the trainee and determine necessary adaptations to the scenario. This paper investigates the automation of didactical knowledge and the corresponding dynamic adaptation of the scenario. A so-called director decides upon and distributes the necessary changes in real-time to the characters. First, the nature and goals of the adaptations are analyzed. Subsequently, the paper introduces a conducted study into the applicability of directable scenarios. Thereafter, an experiment is introduced that investigates the effects of directorial interventions upon the instructive quality of the scenario. Qualitative results indicated that trainees experienced scenario-based training to be instructive and motivating. Moreover, quantitative results showed that instructors rated directed scenarios as significantly better attuned to the trainee's needs compared to non-directed scenarios. Our future research will focus at the design of an architecture for automatically directed scenario-based training.

1. Introduction

Professionals whose future jobs involve the need for complex skills, such as firemen, soldiers, pilots, or nurses, cannot learn their profession from books and lectures alone. Complex skills are comprised of integrated physical and cognitive abilities, e.g., situational assessment and decision making, and require practical training. Research by Kirschner, Sweller, and Clark (2006) showed that in order for practical training to be effective, careful thought should be given to the selection of suitable learning objectives. Moreover, there is a need for online control over the types of learning situations presented to the trainee over the course of training.

Scenario-based training (SBT) is a practical training form. Its goal is the acquisition of complex skills by letting the trainees perform their job in short realistic story lines, called scenarios, which address well-formulated learning objectives (Cannon-Bowers, Burns, Salas, & Pruitt, 1998; Oser, 1999; Salas, Priest, Wilson, & Adler, 2006). To increase the likelihood of the scenario resulting in useful learning situations, scenarios are selected and controlled as the research by Kirschner et al. (2006) suggests. Scenarios are designed and developed by professional instructors within the training domain to fit the learning objectives for each part of the training. Moreover, during training, these instructors make sure the scenario develops as intended by monitoring the course of events as well as the trainee's performance. If the scenario is not developing in a desired direction or if the trainee's performance indicates the need for a change of plans, they decide upon and execute interventions in the scenario where possible.

The adaptation of the scenario ordinarily takes place by altering the behavior of the characters with whom the trainee needs to interact (e.g., teammates, opponents, patients). Therefore, dynamic adaptation of the scenario often requires not just one instructor, but a whole

* Corresponding author. Intelligent Systems Group, Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands. Tel.: +31 88 866 28 85.
E-mail address: m.m.m.peeters@uu.nl (M. Peeters).

0360-1315/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compedu.2013.07.039

team of staff members to play the roles of such characters. Professional SBT instructors use their creativity and expertise to orchestrate the scenario in a desirable direction as the scenario progresses. The scenario does not always unfold the way it was planned. The trainee might, for instance, make some unforeseen choice or the learning situation may turn out to be less suitable for a particular trainee. In such occasions the role players need to improvise and steer the scenario in a (from a didactical perspective) desirable and (from a domain expert perspective) realistic direction. Hence, staff members are required to play these roles.

Even though SBT is a very effective form of training (Cannon-Bowers et al., 1998; Oser, 1999; Salas et al., 2006), the organization of such training opportunities is cumbersome; staff members are not always available and their deployment is usually expensive. This has led to a growing demand by training organizations for forms of training that require less or no human resources (autonomous training). Applications for training purposes to meet this demand have already been developed; however, these applications usually merely simulate the environment and sometimes the characters. In order to exercise the necessary didactical control over the scenario, human instructors are currently still of essential importance, since the requirement for effective control over the training program needs to be satisfied. To truly offer trainees the possibility to engage in training independent of the presence of staff, instructors or teammates, more knowledge is needed about the way instructors control and alter the scenario as it develops. By developing applications that support autonomous training, the number of training opportunities could largely increase, but the question then remains: How can we automate effective didactical control over the scenario?

In this paper we investigate how control over SBT can be performed automatically by adapting the scenario in real time. We will call these adaptations directions, since they give direction, not only to the role players, but to the entire training scenario, hence, also to the trainee. Section 2 describes a theoretical exploration of the ways human instructors conduct their diagnosis and what types of interventions they consider to improve the learning quality of the situation. The theoretical section presents an overview of earlier work relevant for our purposes. Section 3 describes a first study into the creation of directable scenarios. Section 4 discusses an experiment which was conducted to test and evaluate the effects of directions upon the quality of scenario-based training. Finally, the resulting conclusions and implications for future research are presented in Section 5.

2. Automated directions in scenario-based training

The current section starts by providing the reader with some insights in scenario-based training. Subsequently, an argument is presented for the automation of directions in the scenario. This argument is followed by a further refinement of the nature and goals of these directions. Along the lines of reasoning, the envisioned technical realisation of this automation is explained as well.

2.1. Scenario-based training (SBT)

During SBT, trainees participate in story lines, called scenarios, which encompass a causally and temporally related series of learning situations. Each learning situation is linked to a set of learning objectives suitable for the trainee to learn from. To ensure the realism and the didactical value of the scenarios, they are authored and prepared in advance of the training session by an instructor, who is also a subject matter expert. As explained in the introduction, staff members are often required to play the parts of the various characters the trainee needs to interact with.

SBT is a training method directed at the preparation of future professionals to perform complex tasks, such as fire fighting, first aid, military missions, medicine, etc. Complex skill development calls for training methods that (1) integrate all of the constituent skills, knowledge and attitudes, (2) stimulate the trainee to coordinate the constituent skills, and (3) facilitate transfer of the learned skills to new problem situations (Merrill, 2002; Van Merriënboer, 1997). During SBT, this is done by requiring the trainees to prepare, execute and evaluate real-life, relevant and meaningful situations within a simulated environment (SE) (van den Bosch & Riemersma, 2004; Cannon-Bowers et al., 1998; Oser, 1999; Peeters, van den Bosch, Meyer, & Neerincx, 2012; Salas et al., 2006).

Training within an SE offers important benefits, since it can be reused and improved upon by making use of previous experiences. Moreover, it can be controlled, which means that, contrary to training-on-the-job (1) the risks and dangers involved (e.g., the danger of handling weapons or the risk of a patient getting severely ill) can be confined within a safe environment and (2) the complexity of the learning task and the amount of offered support can be gradually increased over the course of training.

To foster transfer, the fidelity of the SE may vary, ranging from the actual task environment to highly symbolic representations thereof. What is important though, is that the SE contains an acceptable level of resemblance to the actual task environment on its most important aspects, such as environmental features necessary for learning and cues that are similar to those found in the actual task environment (Baldwin & Ford, 1988; Young, 1993). Such meaningful, realistic learning tasks are referred to in the literature as "authentic tasks" (Grabinger & Dunlap, 1995; Rieber, 1996). The aspects most important for fidelity purposes vary with the task domain. SEs and scenarios should therefore be developed in consultation with domain experts in order for the SE to contain the necessary resemblances.

2.1.1. Virtual environment

One particular type of SE is the virtual environment. Within a virtual environment (VE), the task environment is simulated by a computer. The use of a VE creates opportunities to implement artificial intelligence techniques, i.e., intelligent agents, that control parts of the environment, such as non-player characters (NPCs) and dynamic environmental aspects (e.g., fires or doors). Examples of agent-based virtual environments for training purposes already exist (van den Bosch, Harbers, Heuvelink, & van Doesburg, 2009). Such systems allow for bigger reductions in costs because they require less organization, less needed man-power, less time to rebuild the environment, and less planning.

Several educational researchers have agreed upon the potential didactical power of such systems (Dickey, 2005; Egenfeldt-Nielsen, 2006; Rieber, 1996). The combination of a virtual environment and intelligent agent techniques can be used to replace the team of staff members by a multi-agent system. However, upon closer look, it becomes obvious that such a system still lacks an automated didactical component. The characters are represented by agents, but they are merely designed to play their parts. As stated in the introduction, control over the scenario requires for monitoring the course of events and the trainee's performance as well as decision making processes about and execution of possible ways to intervene in the scenario as it develops. There is a need for one agent – a director – who is responsible for the didactical edification, while the others only carry the responsibility for their characters' believability. Unlike the other agents, the director does not have control over one particular NPC. Instead, this director is able to instruct the NPCs to change their behavior when necessary. This way the director can steer the course of the scenario into the desired direction, whether for didactical or narrative purposes.

2.2. Didactical directions

During SBT, instructors and role players use their experience and creativity to decide upon interventions during the scenario. Instructors are usually able to recognize situations in which a trainee seems lost, overwhelmed or bored and may decide to adapt the scenario. If they decide to do so, they do this for a variety of reasons, e.g., to adjust the level of support, to let the trainee experience the consequences of his mistakes, to adjust the trainee's self-confidence, but also to maintain a believable storyline, etc.

Domain experts' creativity and experience are often implicit (Dreyfus & Dreyfus, 2005), however, to be able to automate staff members' interventions, we need to turn their implicit knowledge into explicit descriptions, for instance by defining behavioral cues and events that accompany confusion or boredom, e.g., the amount of activity, the amount of mistakes, etc. Subsequently, this automation requires explicit reasoning rules about applicable interventions and the way these can be implemented.

2.2.1. The goal of didactical directions

The previous subsection described how instructors reason and decide upon interventions in the scenario. One of the main reasons for directions is to improve the quality of training, however, the question remains whether and how the quality of training can be increased by executing these directions.

An answer to this question can be found by reasoning about the goal of control over training and creating suitable learning situations. The need for monitoring the trainee before making such decisions points to a relation between the trainee's performance and the decision on whether or not to adjust the learning situation. Apparently the trainee's current performance level serves as a guideline to make decisions about directions. According to Merrill (2002) instruction is an iterative process of supporting the trainee in activating his prior knowledge, demonstrating new knowledge by means of a real world problem, motivating the trainee to apply the new knowledge by himself to a similar problem, and subsequently integrating the new knowledge into everyday life. This shows how support is gradually decreased by starting with a demonstration, followed by stimulating the trainee to practice the problem by himself. However, according to Kirschner et al. (2006) and Van Merriënboer, Kirschner, and Kester (2003) it is important that the trainee receives support during practice as well, again gradually decreasing the amount of support over time. But the question remains, how can one tell what amount of support is sufficient?

When it comes to the matching of the offered training situation to the trainee's skill level, the zone of proximal development (ZPD) comes to mind (Vygotsky, 1978). The ZPD is a controversial concept; researchers are still debating about the intended meaning of the ZPD by Vygotsky (Chaiklin, 2003; Kinginger, 2002). However, the concept is widely used with a variety of meanings and interpretations.

In the current paper, the meaning of the ZPD is regarded to be the set of tasks the trainee is able to perform now with the help of someone (or something) more experienced and independently in the near future (Brown et al., 1993). Offering a trainee learning situations residing within the ZPD is supposed to improve the quality of training, because the trainee is pushed to perform new tasks and gain new experiences and insights. Brown et al. (1993) describe this as follows: "The zone of proximal development embodies a concept of readiness to learn that emphasizes upper levels of competence. Furthermore, these upper boundaries are seen not as immutable but as constantly changing with the learner's increasing independent competence at successive levels." Looking at Fig. 1 (a), it becomes clear how the ZPD can be interpreted as a search for balance between the level of challenge offered to the trainee and the trainee's skill level (Murray & Arroyo, 2003).

This balance between challenge and skill level can also be found in a range of papers published regarding computer game design, except the term frequently used in these publications is flow (Chen, 2007; Rieber, Smith, & Noah, 1998; Shute, Ventura, Bauer, & Zapata-Rivera, 2009; Sweetser & Wyeth, 2005). Flow, see also Fig. 1(b), is defined as focused motivation; a state of optimal experience, where a trainee is so engaged in the learning activity that he loses his sense of time, his self-consciousness disappears, and he feels intrinsically motivated to engage in a complex goal-directed activity, simply for the exhilaration of doing (Csikszentmihalyi, 1991). Characteristics for the circumstances of the learning activity are: direct and immediate feedback, clear goals, a sense of personal control, absorption into the activity (immersion) and a balance between ability level and challenge. Jackson, Thomas, Marsh, and Smethurst (2001) found flow to be significantly related to athletic performance, as well as to a positive self-concept, activation and strategic use of psychological skills. These results also indicate the potential usefulness of this concept to complex skills development.

Interventions that are executed to increase the level of support or challenge in order to help and motivate the trainee could be seen as *didactical directions*. Such didactical directions are executed to match the events in the scenario to the educational needs of the trainee.

Elaborating on adjusting the offered amount of support or challenge, we recognize and take into account two types of didactical directions: *supportive* and *challenging* directions. Supportive directions are needed when the events in the scenario cause the trainee to suddenly be faced with an overly complex situation. In such cases, action and support is required to lead the trainee to a less demanding

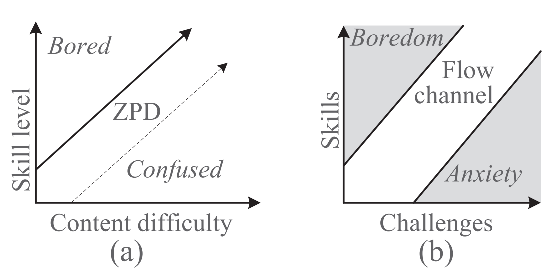

Fig. 1. (a) The zone of proximal development (ZPD) (Murray & Arroyo, 2003) and (b) Flow (Csikszentmihalyi, 1991).

situation. Challenging directions are executed when the trainee is performing all the right actions, but is not being sufficiently challenged and the trainee is motivated to take the training to a higher level. Directions can consist of adjustments of the availability of information, the time constraints, the amount of pressure, the salience of certain cues or the amount of simultaneous actions.

3. Explorative study into the applicability of directions

The models reviewed in the previous paragraph suggest that the training task should be adapted dynamically to the needs of the learner to achieve effective learning and immersion. In their everyday profession, instructors report to apply such regulation, albeit in an implicit manner. However, when thinking about how to exercise automated control over training, more than implicit knowledge is needed. It requires explicit understanding of what cues to monitor when deciding to adapt the course of training. It also requires objective rules on interventions that define precisely when an intervention is appropriate and what it entails.

Furthermore, it is necessary to know more on how learners experience in-between interventions, and whether the interventions indeed improve the quality of the training. Despite the ample attention to adaptation in the literature on learning and training, and its apparent wide-spread use by instructors in practical training programs, very little is known on how automated online direction of scenario-based training can be realized and what its effects are.

In order to learn more about direction of training, an explorative study was conducted. For this study, a four-scenario training program was developed for employees of an in-company emergency management team. For each of the scenarios we determined behavioral cues to monitor, and developed rules for whether or not to apply interventions. Participants received two scenarios in which interventions were applied when cues and rules marked this as appropriate; the other two scenarios were administered without any interventions. The first goal of the study was to learn whether it is possible to define appropriate cues and intervention rules. The second goal was to see whether these cues and rules are feasible to work with; how online intervention is perceived by trainees, and whether the educational objective of a particular scenario is helped, undermined, or unaffected by online interventions.

3.1. Task domain

The task domain for this study was 'In-Company Emergency Management' (ICEM). ICEM entails the application of first aid, fire fighting, and evacuation procedures, all of which are performed by a trained team of company employees. They remain in charge until the incident is sufficiently handled, or when the official emergency services arrive (i.e., ambulance, firemen, police).

3.2. Method

3.2.1. Participants/trainees

All members of an ICEM team (7 males; 3 females) of a research organization company in the Netherlands participated in the experiment. The participants were uninformed with respect to the research questions. Each participant individually took part in the scenarios as a trainee.

3.2.2. Scenarios

In this study, three people were involved in four separate scenarios: the trainee (participant) and two actors. A female actor played the role of victim and a male actor the role of bystander (e.g., colleague). The experimenter fulfilled the role of director, consisting of monitoring the training from an adjacent room by means of a video-connection and sending instructions to the actors through in-ear portophones. Two scenarios involved fire incidents, whereas the other two involved medical emergencies (see Table 1). The scenarios were developed in consultation with an experienced ICEM trainer according to the following steps:

a. Defining learning goals and story lines

A set of selected learning goals formed the foundation of the story lines. An appropriate learning situation was created for each of the selected learning goals. Subsequently, the resulting learning situations were connected to form four coherent story lines.

b. Actor scripts

For each scenario, scripts were constructed that contained explicit instructions on how both actors should behave. The actors received two versions of each of their scripts: a supportive and a challenging version. The supportive script version instructed each actor how to be helpful to the trainee, e.g., by acting calm and stable, offering help, asking guiding questions, or even suggesting possible solutions. In contrast, the challenging script version instructed each actor how to make the situation more difficult, e.g., by acting panicky, being passive, or creating extra complications (for instance, running into a fire hazard).

Table 1
The four scenarios developed varied on the addressed type of the incident, descriptions of the scenarios are provided below.

Name	Type	Description
A	First aid	A diabetic woman suffering from hypoglycemia
B	Fire	A lady trapped within a room because of a small fire in a trash can near the door
C	First aid	An unconscious cleaning lady, who fainted because of an intoxicating gas
D	Fire	A woman with a broken hip – as a result of fleeing in panic from a fire – lying near a fire hazard

It is noteworthy that the learning situations were the same for all participants. The behavior of the actors, however, varied. For example, in the supportive version of the 'diabetic patient' scenario, the trainee spontaneously received relevant information, whereas in the challenging version this information was only provided upon request.

c. Director script

For each scenario, a director script was constructed. The goal of this script was to objectively specify to the director which trainee behavior required which intervention, and at what point in time it should be executed. If the trainee behavior called for an intervention according to the script, the director instructed one or both of the actors through their in-ear portophones to switch to the other version of the script (i.e., from challenging to supportive or vice versa).

The director scripts dictated a shift from supportive to challenging behavior if the situation proved to be too simple for the trainee. The script contained behavior descriptions to determine if the intervention was apposite, for example: 'trainee makes eye contact with victim', 'trainee remains calm', or 'trainee gives clear instructions to bystander'. These behavioral cues would result in the director's intervention to instruct the actors to change their behavior to the challenging version of the script, examples of which are described above in the actor scripts paragraph.

Likewise, the director scripts dictated a shift from challenging to supportive behavior if the situation proved to be too difficult for the trainee's competency level. The script contained behavior descriptions indicating a too difficult learning situation, such as the trainee neglecting victim and bystander altogether, failing to perform all required actions, or performing inappropriate actions. The corresponding instruction to the actors would be to switch their behavior to the supportive version of the script, as described in the actor scripts paragraph above.

3.2.3. Procedure

Each participant was introduced to the actors before the training. Subsequently, they received explanations on their participation in a scenario-based ICEM training in a regular office-setting. Despite the obvious lack of realism, the participants were asked to act as if the incidents were real. They received information about the simulated environment (e.g., locations of the first-aid kit and fire extinguishers) and how to phone the receptionist who, in turn, could call up the official emergency services. Before the start of each scenario, the participant was asked to wait further down the hallway and the actors took their positions. The experimenter then went into an adjacent room and gave permission for the scenario kickoff: the female victim would start screaming for help. The experimenter/director followed the events through a video-connection. Each time the situation and script called for an intervention, the director issued the intervention by instructing the actor(s) through their in-ear portophones. As would be expected, the participants were oblivious with respect to any of the issued interventions. After completing each of the four scenarios each participant was interviewed about their experiences (decisions, insights, emotions, work load, motivation, etc.). After all four scenarios had finished, the participant was interviewed to evaluate his (or her) overall reflections. All interviews were sound-recorded. All training sessions were video-taped using three cameras.

3.2.4. Experimental design

Effects of interventions during training were investigated within-subjects by applying interventions when appropriate in half of the scenarios. In the other half of the scenarios, no interventions were applied (irrespective of their applicability). For each participant, either the first two administered scenarios were directed or the last two. As explained in Subsection 3.2.2, each scenario had two versions (a challenging version and a supportive version), and the scenario could thus be started in either mode. Half of the participants started all four scenarios in the supportive version; the other half of the participants always started in the challenging version. In addition, half of the participants participated in the scenarios in reverse order. The order of the scenarios, the order of directed vs non-directed training, and the starting mode of the scenario were counterbalanced over participants and conditions. This means that there were 8 groups in the study, depicted in Table 2. Participants were randomly assigned to the groups. Pictures of the resulting scenario plays can be found in Fig. 2.

3.3. Results

The design and development of adaptive scenarios included two major aspects. The first aspect was the definition of measurable and explicit behavioral cues to decide whether an intervention was applicable. The second aspect was the definition of fixed adaptations, referred to as interventions, in the scenario.

Identifying measurable and explicit behavioral cues that indicate the need for a scenario adaptation, which was done in consultation with domain experts, was fairly straightforward. Moreover, the resulting cues were unambiguous and traceable. The same applies to the design of the intervention rules. No problems were encountered in recognizing the applicable rule and delegating the corresponding shift to the

Table 2
In order to ascertain a counterbalanced design, participants were randomly assigned to 8 groups, thereby counterbalancing the order of the four scenarios, the order of directed vs non-directed scenarios, and the starting mode (supportive vs challenging).

Group	Scenario order	Initial mode	First two scenarios	Last two scenarios	Participants (n)
1	ABCD	Supportive	Directed	Non-directed	1
2	ABCD	Supportive	Non-directed	Directed	1
3	ABCD	Challenging	Directed	Non-directed	1
4	ABCD	Challenging	Non-directed	Directed	2
5	DCBA	Supportive	Directed	Non-directed	1
6	DCBA	Supportive	Non-directed	Directed	2
7	DCBA	Challenging	Directed	Non-directed	1
8	DCBA	Challenging	Non-directed	Directed	1

Fig. 2. a) The director and the telephone operator watching the training session on the screen in the control room. b) The trainee addressing and diagnosing the victim in the training environment.

actors. By delegating the details of the shift to the actors, the director was able to focus on the behavioral cues and the decisions on rule applicability. In addition, most interventions affected the course of the scenario as intended.

The scenarios and the learning environment were open in nature. Trainees had a large degree of freedom in their behavior. As a result, there was a high variability in the way trainees proceeded through the scenario, with some trainees performing completely different sets of actions than others.

Results from the interviews show that trainees did not experience any differences between directed scenarios and non-directed scenarios. None of the participants reported being aware that the course of events was (in some cases) altered by the director.

Given the exposure to only two scenarios in directed training vs two scenarios in non-directed training, it would be unrealistic to expect that this would result in substantial differences in the quality of learning. Furthermore, this would require clear-cut and objective performance measures for this task and domain, which are currently unavailable. Therefore, the scale, scope and design of this explorative study were not sufficient to draw any conclusions about the effects of interventions on the trainee's ability to achieve the learning objectives of the scenario.

Questions concerning work load, motivation, self-efficacy and emotional state were included in the post-scenario interviews to investigate any possible effects of the interventions on the trainee's personal experiences. Unfortunately, the collected data did not allow for any definite conclusions for two reasons. First, a high variability in the way trainees proceeded through the scenarios, which caused the data to refer to highly different training situations. This variability rendered it impossible to unequivocally attribute results to the effects of direction. Second, even if a participant was performing a scenario in the directed condition, this did not necessarily imply that the interventions were actually issued. For example, if the scenario started in the challenging mode, and the trainee performed adequately, then a shift to the supportive version would never be issued.

Even though the experimenter and assistants were aware of possible biases in their experiences, they still reported perceiving an effect of the interventions on the development of the training scenario. They indicated that when an intervention was issued, the training appeared to be more in line with the individual trainee's competencies, whereas when an intervention would be appropriate but was not issued, the trainee was believed to be led astray more often.

3.4. Discussion

The goal of this explorative study was to extend our knowledge about directed training. In particular, this study investigated 1) whether it is possible to define appropriate cues and intervention rules. In addition, it researched 2) how online interventions are perceived by participants. Moreover, it examined 3) whether the educational objective of a particular scenario is helped, undermined, or unaffected by online interventions.

First of all, the automation of directed training requires objective criteria to decide on the necessity of intervening. In the preparations of this study we consulted a domain expert to define such objective cues and rules. Defining behavioral cues that indicate the correspondence of the training scenario with the participant's learning objectives turned out to be relatively facile. This study shows that properly specified sets of cues and rules allow for objective direction, which is a principle requirement for automated direction of training.

To keep conditions constant, the same sets of cues and rules were perpetuated throughout the entire study. However, observations during the study could have been used to improve and refine the cues-and-rules sets. For use in practical settings, it is recommended to profit from such experiences, and iteratively refine the intervention cues and rules.

Secondly, the scenarios proved to be quite taxing for the majority of the participants. All participants mentioned their involvement and immersion in the scenarios, which might explain their general unawareness of the, possibly, issued interventions. This unawareness hampered the investigation of their experiences regarding the interventions. Moreover, due to differences in competency among the participants, along with high degrees of freedom during training, there was a large variation in the way participants proceeded throughout the scenarios. This variety rendered it almost impossible to examine the effects of directions on the results coming from the interviews. This study therefore does not provide any information on how participants experience online interventions.

Thirdly, anecdotal evidence suggests that interventions result in a higher correspondence between the encountered training situations and the participant's competencies, and that refraining from intervention results in a less-productive learning situation. Possibly, the quality of a scenario is positively influenced by adding directions. Directions may restore the effectiveness of scenarios that would otherwise lose their suitability. In other words, trainees may benefit from directions, even if they are unaware of it. The video recordings obtained during this study offer possibilities to further investigate this hypothesis. They allow for careful selection of the fragments showing a cue for an intervention, since interventions are perfectly traceable because of the scripts. The objective of the interventions is to restore and/or to maintain a learning situation that enables the participant to learn from. Since these fragments show occasions in which the intervention was actually issued (the directed conditions), as well as occasions in which the intervention was not issued (the non-directed conditions), the quality of the emerging learning situations can be compared to address the effects of the interventions. To further examine this issue, a new study was conducted. This study is described in the following section.

4. Evaluating the learning effects of directions

Section 3 described an explorative study into the applicability of a director agent. This director agent executed interventions in the scenario by applying rules if they were applicable according to explicit behavioral cues. The results of the study showed that this design enabled the desired adaptivity of the training scenarios. The director was able to intervene and alter the course of events by instructing the characters in the scenario to change their behaviors based upon these scripted rules. However, the desired effects of these interventions upon the quality of the training scenario still needed to be investigated. The reason for intervening in the scenario, was to improve the quality of training by adjusting the level of support to match the trainee's level of performance. In order to test these effects, 'quality of training' needs to be further specified.

In Section 2.2.1 it was argued that a learning situation offers optimal learning opportunities if a trainee is able to cope with the demands, while still being challenged to learn new things (Murray & Arroyo, 2003). The proximity of a training situation to this optimum can be expressed as the *learning value*. If a training situation has a low learning value, this means that the situation does not meet the trainee's needs: the trainee is either incapable of coping with the demands or he is not being challenged enough to be motivated. In both cases an intervention would be necessary to attune the scenario to the trainee's needs.

To investigate the effects of directions upon the learning value of the scenario, an experiment was conducted. The aim of the experiment was to answer the question: Do the interventions in the study described in Section 3 actually lead to an improvement of the learning value?

To answer this question, the video material coming from the study described in Section 3 was analyzed. We traced back the fragments containing the need for a directorial intervention, as defined by the director's script (see also Section 3.2.2). Such a directorial intervention might or might not have been executed, depending on the presence or absence of the director. The resulting fragments were then judged by experts with respect to the learning value of the learning situation. Their judgments were then analyzed to investigate the effects of the interventions on the learning value of the scenarios.

4.1. Research question

The research question that is to be answered by means of this experiment is: *Will the interventions of a director during scenario-based training improve the learning value of the training scenario according to domain experts?*

We hypothesize that interventions of a director improve the learning value of the training scenario as rated by professional instructors.

4.2. Methods

4.2.1. Raters

Six experienced instructors in ICEM (in-company emergency assistance, see also Section 3.1) were asked to rate the video-fragments.

4.2.2. Materials

4.2.2.1. Footage. For this experiment we selected twenty video fragments as a test set. We will provide some extra information on how this selection took place below. Each fragment contained a part of a recording displaying a trainee as he played one of the aforementioned ICEM scenarios. All selected video fragments contained trainee behavior cueing an intervention. In half of the fragments shown to the instructors, the director executed all expedient interventions (*directed condition*) by telling the actors through in-ear portophones to switch between their behavior variations. In the other half of the shown fragments, the director was absent and the cues in the protocol were ignored (*non-directed condition*). No interventions were executed in this condition. Additionally, both conditions (directed and non-directed) contained five fragments that started off with the actors playing their supportive parts (*supportive start up*), and five fragments that started off with the actors playing their challenging parts (*challenging start up*).

4.2.2.2. Selection of the fragments. The nature of the repeated measures analysis requires for an equal distribution of the amount of fragments over the different conditions. This means that the maximum number of fragments for each condition is the number of fragments in the condition containing the smallest sample. In our case the smallest condition sample contained five fragments. Based upon this amount, we had to remove fragments from the other conditions to leave five in each condition. This was done based upon careful selection. We will now give the reasons for removal of fragments ordered by applicability:

- The intervention was executed shortly after the scenario had started, therefore the raters would not have much time to get a clear idea of the skill level of the participant.
- The fragment contained a lot of action outside of the scope of the camera.
- The intervention was not purely pedagogical; the purpose was also to maintain the believability of the storyline.
- One of the actors started improvising, not sticking to the script as planned.

The resulting selection of twenty fragments contained only fragments that were representative for their condition and contained enough time before the intervention for the instructors to get a clear idea of the development of the scenario.

4.2.2.3. Questionnaire. The raters were asked to evaluate the learning value of the situation for a particular trainee by answering the following question:

"The learning situation at this point in time offers the trainee …opportunities to achieve the learning goals at his own level"

-3	−2	−1	0	1	2	3
absolutely no	no	not really any	maybe some	some	enough	outstanding

4.2.3. Procedure

The raters received an elaborate instruction to this experiment. The instruction contained an explanation of scenario-based training, exemplified by a video fragment. The four scenarios were explained and the learning goals for each scenario were explicitly pointed out. Finally the raters received instructions regarding the procedure of the experiment and explanations to the questionnaire. The raters were oblivious of the research question of the experiment.

Raters were then presented with two sets of video fragments (a practice set and a test set) following a standard procedure. The video fragment was introduced by a short description of the original scenario and the intended learning goal. The part of the fragment preceding the point of intervention was shown. At the cue for intervention, the fragment was paused and the raters were asked to rate the learning value (rating moment 1). Subsequently, the fragment was continued and paused again at the time the result of the intervention (or the lack thereof) became apparent. The raters were again asked to rate the learning value (rating moment 2). A diagram of the procedure can be found in Fig. 3.

To test and enhance agreement among raters, they were presented with a practice set of 16 video fragments. The raters were encouraged to discuss their judgments in between series to reach consensus on how to value a learning situation. After the practice set, the experiment proper started, by presenting the test set consisting of twenty video fragments to the raters. The raters were not allowed to discuss their judgments, nor could they see each other's judgments. After the test set, the raters participated in a group discussion about their experiences with scenario-based training and their opinions about the video fragments.

4.2.4. Analysis

An intra-class correlation analysis was performed to assess inter-rater reliability. A repeated measures ANOVA was used to compute the effects of direction upon the rated learning value of the scenario.

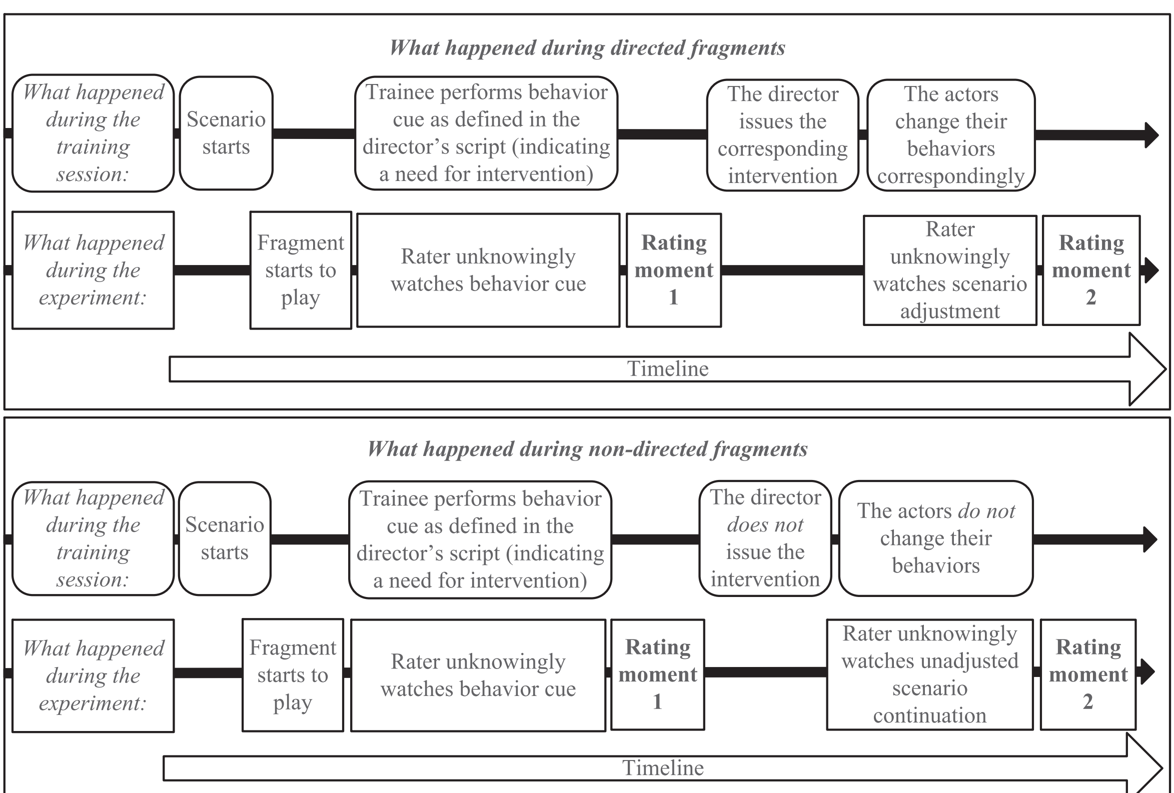

Fig. 3. A graph of the procedure during the experiment and its relation to the previous study.

Table 3
results of the repeated measures analysis.

Effect	Director[a]	Start up variation	Director × rating moment[a]
F	13.847[b]	11.043[c]	27.339[b]
Effect	.735	.688	.845
Power	.841	.757	.984

[a] One-tailed.
[b] $p<.01$.
[c] $p<.05$.

4.3. Results

4.3.1. Data exploration and inter-rater reliability

Forty ratings per rater (two rating moments for a total of twenty fragments) were entered into the analysis. The consistency intra-class correlation coefficient was .694 for average measures ($p < .001$). An inter-rater agreement between .60 and .79 is considered substantial (Landis & Koch, 1977), therefore we consider these data to be appropriate for further analysis.

4.3.2. Repeated measures analysis

In order to test whether the interventions of the director had an effect on learning value, rated learning values were entered into a repeated measures analysis with two independent factors: director (presence vs absence) and start up variation (a scenario starting in the supportive vs challenging behavior variation). The results of this analysis are shown in Table 3.

A main effect of direction was found ($F(1,5) = 13.85$; $p < .01$, one-tailed). Examination of this effect showed that the directed fragments received a significantly higher learning value ($M = 1.08$; SE = .31) than the non-directed fragments ($M = .35$; SE = .23). A second main effect showed a significant difference between the learning value assigned to the two start up conditions ($F(1,5) = 11.04$; $p < .01$, two-tailed). Overall, the video fragments in the supportive start up condition received a higher learning value ($M = .98$; SE = .31) than those in the challenging start up condition ($M = .45$; SE = .22).

Our main interest is the effect of an intervention on the situation's learning value. Therefore the differences between the director conditions (present vs absent) at rating moment 2 are of importance. It is expected there are no differences between the two conditions at rating moment 1. A significant interaction effect between director (presence vs absence) and rating moment (prior to vs after the cue for intervention) ($F(1,5) = 27.34$; $p < .01$, one-tailed test), showed that indeed there was no significant difference between the directed and the non-directed condition at rating moment 1 ($M = .60$ vs $M = .43$, respectively). However, if an intervention was executed at rating moment 2 (director present), the learning value was significantly higher than when no intervention had taken place (director absent) ($M = 1.55$ vs $M = .27$, respectively). The means belonging to this interaction effect can be found in the row 'overall' of Table 4.

To find out whether the beneficial effect of the director's interventions is equal for both directions of interventions (from supportive to challenging or vice versa), one-tailed 95% confidence intervals of the means were computed for both start up conditions. The interaction effects were significant ($p < .05$, one-tailed) for both directions of intervention, (see also Table 4), although the effect was stronger for supportive interventions (changing the actor behavior from challenging to supportive).

4.4. Discussion

The goal of this experiment was to investigate the effects of interventions upon the learning quality of a scenario. We created scripts for a director specifying when and how to intervene. Interventions consisted of adaptations in the behavior of the actors (NPCs) and were implemented online, i.e., while the scenario unfolded. Video recordings of directed and non-directed training scenarios were shown to experienced instructors, who were asked to rate the learning value of the presented situations. Instructors were naive with respect to the purpose and design of the experiment.

Results confirmed our hypothesis. The rated learning value of scenarios that proceed non-directed, without adaptation, were at a fairly low level both halfway and at the end of the scenario. In contrast, the learning quality of directed scenarios improved significantly as a result of the interventions directing the actors to behave appropriately to the performance level of the trainee. Thus, overall, interventions improve the learning value of scenarios. If we examine these results more closely, split for supportive and challenging start up conditions, it becomes clear that scenarios that started in the supportive mode also offer some learning opportunities in the absence of a director. Even though the trainee could use an extra challenge, the mere practice of already acquired skills is still considered useful. However, in the directed condition, it becomes possible to create an extra challenge for the trainee, which results in an even higher learning value. A different pattern is found for the scenarios that started in the challenging mode. For these scenarios, the learning value drops dramatically over time when there is no director present to adjust the scenario. However, in the presence of the director, support is given to the trainee, thereby most likely saving the trainee from losing track and motivation and increasing the learning value of the training.

In a group interview conducted after the experiment, we explained the purpose and design of the study to the instructors and asked them for their experiences in their everyday work. The instructors stated that they find it hard to successfully intervene once they notice that a scenario loses track. They argue that they do realize it when a training situation requires intervention, but that they find it hard to specify beforehand what cues indicate this need. A more practical problem that they put forward is that – in their experience – participating actors tend to be unaware of what is needed, and that it is difficult for instructors to bring across appropriate adjustments to the actors while the scenario is playing. Instructors therefore consider it important to have appropriate and practical training instruments to execute the necessary control over their training scenarios.

The developed rules, consisting of explicitly described cues that refer to trainees' responses, which trigger different types of interventions in a training scenario, showed to be not only highly desired by experts working within the training domain, but to be beneficial for the learning value of the scenario.

Table 4
Mean rated learning value (SE).

	Director present		Director absent	
	Moment 1	Moment 2	Moment 1	Moment 2
Challenging start-up	.433 (.336)	1.467[a] (.470)	.233 (.285)	−.333 (.276)
Supportive start-up	.767 (.391)	1.633[a] (.363)	.633 (.336)	.867 (.418)
Overall	.600 (.306)	1.550[a] (.394)	.433 (.262)	.267 (.324)

[a] $p<.05$, one-tailed.

5. Conclusion

This paper elaborates on the concept of a director for scenario-based training. Such a director is able to automatically manipulate the scenario in real time to create correspondence between the scenario's offered learning situations and the trainee's skill level. The design of this type of manipulation requires a careful analysis and formulation of predetermined rules enabling the director to intervene in the scenario. Interventions in the scenario are meant to adjust the level of challenge and support. Our approach was to create predetermined intervention rules, consisting of a specific behavior displayed by the trainee (conditional), which triggered a shift in the behaviors displayed by the characters in the scenario (consequence). By alternating the behavior of the actors between a challenging and a supportive version of their parts, the scenario could be adjusted in the amount of support or challenge offered to the trainee.

Section 3 reported on a study into the applicability of this concept. Four adaptive scenarios were developed within the domain of in-company emergency assistance (see Subsection 3.1). This study showed that, even though the directions in the approach described here were simple scripted rules, they still led to variable, dynamic, and interactive scenarios. Moreover, it revealed a challenge for subjective measures; the high variability makes comparisons between the scenarios and the subjects an intriguing problem for future experiments, which demands for creative experimental design and measurements.

Section 4 describes an experiment in which we investigated the effects of a director's interventions upon the quality of learning. The results of this experiment showed that instructors judge the director's interventions in the scenario to induce learning situations with a significantly higher learning value, in comparison with situations in which these interventions were not executed.

Based upon this research we conclude that the approach of a director agent is promising. The realism and authenticity of the job are warranted, whereas simultaneously, control over the course of training is maintained. In addition, the design approach is intuitive and straightforward. Moreover, it offers automated control over the scenario in a way that instructors and staff-members would want to influence the training scenario in order for the trainees to learn from it.

5.1. Future work

In the current paper the scenario adaptivity consisted solely of online adaptations. The learning goals were defined offline and adaptations were designed to adjust the amount of offered support and challenge to the trainee. Future work will focus on a further elaboration of the functionalities of the director, e.g., scenario selection, learning goal adaptation, curriculum planning and feedback provision. Identification of the desired functionalities involves a requirements analysis of the director agent. This entails consultation with subject-matter experts, performing additional literature research and conducting repetitive research cycles to refine the requirements of automatically directed scenario-based training. The constructed requirements baseline will lead to a modular design for the director agent architecture. Prototype development will then enable further validation and evaluation of the requirements and architecture. By continuing these research cycles in the future the director agent architecture can be further refined and implemented.

References

Baldwin, T., & Ford, J. (1988). Transfer of training: a review and directions for future research. *Personnel Psychology, 41*, 63–105.

van den Bosch, K., Harbers, M., Heuvelink, A., & van Doesburg, W. (2009). Intelligent agents for training on-board fire fighting. In *Proceedings of the 2nd international conference on digital human modeling* (pp. 463–472). Springer.

van den Bosch, K., & Riemersma, J. (2004). Reflections on scenario-based training in tactical command. In S. Schiflett (Ed.), *Scaled worlds: Development, validation, and applications* (pp. 1–21). Surrey, UK: Ashgate Publishing Ltd.

Brown, A., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., & Campione, J. (1993). Distributed expertise in the classroom. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 188–228). Cambridge, UK: Cambridge University Press.

Cannon-Bowers, J., Burns, J., Salas, E., & Pruitt, J. (1998). Advanced technology in scenario-based training. In J. Cannon-Bowers, & E. Salas (Eds.), *Making decisions under stress* (pp. 365–374). American Psychological Association.

Chaiklin, S. (2003). *The zone of proximal development in Vygotsky's analysis of learning and instruction*. In *Vygotsky's educational theory in cultural context* (pp. 39–64).

Chen, J. (2007). Flow in games (and everything else). *Communications of the Association for Computing Machinery, 50*, 31–34.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience: Steps toward enhancing the quality of life*. Harper Collins Publishers.

Dickey, M. (2005). Engaging by design: how engagement strategies in popular computer and video games can inform instructional design. *Educational Technology Research and Development, 53*, 67–83.

Dreyfus, H., & Dreyfus, S. (2005). Peripheral vision: expertise in real world contexts. *Organization Studies, 26*, 779–792.

Egenfeldt-Nielsen, S. (2006). Overview of research on the educational use of video games. *Digital Kompetanse, 1*, 184–213.

Grabinger, R., & Dunlap, J. (1995). Rich environments for active learning: a definition. *Association for Learning Technology Journal, 3*, 5–34.

Jackson, S., Thomas, P., Marsh, H., & Smethurst, C. (2001). Relationships between flow, self-concept, psychological skills, and performance. *Journal of Applied Sport Psychology, 13*, 129–153.

Kinginger, C. (2002). Defining the zone of proximal development in us foreign language education. *Applied Linguistics, 23*, 240–261.

Kirschner, P., Sweller, J., & Clark, R. (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.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development, 50*, 43–59.

Murray, T., & Arroyo, I. (2003). Toward an operational definition of the zone of proximal development for adaptive instructional software. In *Proceedings of cognitive science. Boston, MA*.

Oser, R. (1999). A structured approach for scenario-based training. In *Proceedings of the 43rd human factors and ergonomics society annual meeting annual meeting* (Vol. 43; pp. 1138–1142). SAGE Publications.
Peeters, M., van den Bosch, K., Meyer, J., & Neerincx, M. (2012). Situated cognitive engineering: the requirements and design of automatically directed scenario-based training. In *Achi 2012, the fifth international conference on advances in computer-human interactions* (pp. 266–272).
Rieber, L. (1996). Seriously considering play: designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research & Development, 44*, 43–58.
Rieber, L., Smith, L., & Noah, D. (1998). The value of serious play. *Educational Technology, 38*, 29–36.
Salas, E., Priest, H. A., Wilson, K. A., & Adler, A. B. (2006). Scenario-based training: improving military mission performance and adaptability. In T. Britt, A. Adler, & C. Castro (Eds.), *Military life: The psychology of serving in peace and combat* (pp. 32–53). Praeger Publishers.
Shute, V., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). *Melding the power of serious games and embedded assessment to monitor and foster learning*. In *Serious games: Mechanisms and effects* (pp. 295–321).
Sweetser, P., & Wyeth, P. (2005). Gameflow: a model for evaluating player enjoyment in games. *Computers in Entertainment, 3*, 1–24.
Van Merriënboer, J. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Educational Technology Publications.
Van Merriënboer, J., Kirschner, P., & Kester, L. (2003). Taking the load off a learner's mind: instructional design for complex learning. *Educational Psychologist, 38*, 5–13.
Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
Young, M. (1993). Instructional design for situated learning. *Educational Technology Research and Development, 41*, 43–58.

Marieke Peeters (Ma. Sc.) is a PhD-student at the department of computer science at Utrecht University. She combines her knowledge about instructional theory and artificial intelligence to investigate adaptivity in serious games by manipulating the storyline to match it to the trainee's knowledge and skill set.

Dr. Karel Van Den Bosch is senior researcher at TNO. His work involves the development and deployment of intelligent software agents in simulation- or game-based training. The goal is to make training more effective (by consistently eliciting the intended behavior of the trainee), and more efficient (e.g., team members no longer need to be present).

Prof. Dr. John-Jules Ch. Meyer is a full professor of computer science at Utrecht University where he heads the Intelligent Systems Group. His current research interests include artificial intelligence, and intelligent agents in particular, from theory to applications. He is a Fellow of the European Coordinating Committee for Artificial Intelligence.

Mark Neerincx is professor in Human–Machine Interaction (HMI) at the Delft University of Technology, and senior researcher at TNO Perceptual and Cognitive Systems. He has extensive experience in fundamental and applied research on human–computer interaction, among other things in the domains of health, security, defense, and space. Important results are (1) cognitive task load and emotion models for performance and health support, (2) models of human-machine partnership for attuning assistance to the individual user and momentary usage context, (3) prototypes of electronic partners for self-efficacy, (4) cognitive engineering methods and tools, and (5) a diverse set of usability "best practices".