Original article

doi: 10.1111/jcal.12105

Journal of Computer Assisted Learning

Assessment based on Serious Gaming Interactive Questions (SGIQ)

G. Šimić,* A. Jevremović,† Z. Kostić‡ & D. Đorđević§

*University of Defense, str. Generala Pavla Jurisica Sturma 33, 11000 Belgrade, Serbia
†Sinergia University, str. Raje Banjičića, Bjeljina 76300, Bosnia and Herzegovina
‡Singidunum University Belgrade, str. Kumodraska 261, 11000 Belgrade, Serbia
§Military Medical Academy, Crnotravska 17, 11000 Belgrade, Serbia

Abstract The case study presented in this paper describes the pedagogical aspects and collected experience in using e-learning tool named IPA-PBL. Improving assessments in the preparation for AMET's (Air Medical Evacuation and Transport) complex task of transfer of injured or sick patients from the place of accident to the hospital or between hospitals represents the main motive for developing this tool. In IPA-PBL, the problem-based learning is applied as a pedagogical strategy as well as a set of concrete methods implemented in the software. Together with the pedagogical model, specific domain ontology is designed. This way, the learner's knowledge can be analysed in order to collect the data necessary for dynamical adaptation of system behaviour. The results collected during the exploitation of IPA-PBL are compared without using the system. Statistical analysis as well as considerations and conclusions about them are also presented in the paper. The predominant number of trainees who participated in the survey expressed positive attitudes about the new interactive type of questions. Therefore, it is considered to extend using such questions not only in assessments but in other parts of the training process.

Keywords assessment systems, interactive 3D virtual environments, serious gaming.

Introduction

This paper presents the results of design and using the new question type – Serious Gaming Interactive Questions (SGIQ). The main goal of the development of SGIQ is to improve the assessments by using highly interactive virtual space as a part of questions in order to incorporate space model into the tasks, to track the students' behaviours in a new way and to motivate them for achieving better results than before. It is produced by using X3D interactive technology (http://www.web3d.org/) embedded in the question content delivered as a Web form and used in assessments with specific purpose. The domain context for which the SGIQ is primarily developed is medical aid – preparing the medical staff for special missions in very constrained conditions. The doctors, surgeons and their assistants represent the target group. Regardless of their skills and expert domain knowledge, they have no experience in giving the medical support in land and aircrafts, during the transport and by using reduced equipment. Therefore, the main focus was on creating the new assessment environment which can improve their motivation. The design of test sets for high-skilled adults who already have sufficient domain knowledge and lots of references in the previous work represented the challenge for designers and required sophisticated

Accepted: 21 December 2014
Correspondence: Zona Kostić, Singidunum University Belgrade, str. Kumodraska 261, 11000 Belgrade, Serbia. Email: zonakostic@gmail.com

approach in order to attract them to make more than one attempt for achieving better results during their preparations for missions.

Contemporary educational systems as well as business systems use the IT technologies in every segment of education and training (E&T). The organizations find these technologies appropriate for performing their activities in more effective way than before. They spare both the money and time. Therefore, the interactive e-books and/or e-courses are becoming more used as supplements or substitutes for hard-printed materials. In the same way, the simulations and serious gaming are more used than the live exercises while the paper test sheets are replaced by assessments performed by using computers. Sometimes it is forgotten that the effectiveness is not the main goal of E&T. For instance, there are lots of examples where the institutions use the robust learning management systems only for delivering the non-interactive content to their users. Many of other functions offered are not used at all.

The assessments represent the most sensible part of using technology in E&T. Making meaningful interactive assessments still represents a big challenge for the teachers. It is not unusual that they deliver the responsibility for creating the tests to their assistants due to number of students, courses and exams as well as their engagements in research projects and administrative work. Without the control, feedback and detailed analysis of student results, the assessments could direct the students' motivation in an undesired way. If the 'true/false' questions are the only type used, it is likely the students will pass the exam only based on recognition. This type is the simplest one and as the results become unexpectedly high the usual teachers' reaction is to redesign 'true/false' questions into the 'multiple choice' type. It improves the assessments but the students' success is still based on recognition level of knowledge. Unfortunately, the further development of the test questions needs much more engagement of both the teachers and their assistants, and therefore the 'multiple choice' tests often become a dominant type of e-assessments used. While the students are always concerned about passing the exams with as few preparations as possible, regardless of the nice results they achieve, their real skills and knowledge are usually at unsatisfactory level which represents the real damage of poorly designed assessments. It represents one of the motivation factors for developing the new type of questions that can improve the trainees' real knowledge and their motivation to achieve the best results they can.

The case study presented is about assessments in the preparation for AMET's (Air Medical Evacuation and Transport) complex task of transfer of (seriously wounded or sick) patients from the place of accident to the hospital or between hospitals of different levels (http://www.vs.rs/index.php?content=4423c1a4-56bb-102f-8d2f-000c29270931#MONUSCO). This task has been performing since 1999 as a regular task of UN (United Nations) missions in Congo, and Serbian medical team has participated in it since 2003. The standard equipment and procedures are established based on Warren *et al*. (2004). The need for development of a new kind of assessment was recognized in previous research (Šimić, 2012). Besides the expert knowledge, the medical staff engaged in the mission need to follow these procedures during the transfer regardless of environmental conditions that could be very hard and sometimes stressful. Moreover, the medical staff need to provide almost all kinds of urgent interventions and critical health care in very limited conditions (small room, with or without few assistants, portable equipment, changeable transport conditions, etc.). For instance, the anaesthesiologists were in situations to use manual respirators to keep breathing or to give intravenous therapy and blood transfusion. Sometimes, they need to deal with patients in the land and aircrafts they have never been before. They pass through some training, but the situation on the terrain is changeable. Therefore, they need to know where the equipment is positioned in the craft as well as to be familiar with different types of equipment before sending on a mission. In addition, physicians and medical technicians should know the critical care procedures that have to be conducted during the transport.

The science foundation used in the research is described briefly in the next section. The third section is focused on technology aspects of the developed system. The application logic is split in client and server sides and the specific solution is developed for data exchange between them. There are three types of SGIQ tasks developed and one part of the third section is dedicated to their descriptions. Evaluation of the system and considerations of trainees' results are described in the section 'After assessment analysis'. Finally, the conclusions and future plans are presented in the last section.

Related works

In the assessment design, the main effort was in providing high interactivity and realistic conditions during the examination. According to the problem presented and the assessment objectives described in previous section, it is obvious that it is expensive and often impossible to provide exactly the same equipment and aircrafts that are used in the mission. Therefore, the virtual working space had to be designed. The technologies needed have been already used in serious gaming. As described in Zyda (2005), in serious gaming the entertainment is used for educational purposes while pedagogy is used for adapting the game scenario according to the educational objectives. In the presented case study, the serious game concepts are used for improving the quality of assessments for special mission with precisely defined goals. It is obtained by developing the assessment scenarios in virtual environments as much as possible similar to realistic ones. They are also used to motivate the trainees to achieve better results than before.

Web-based visualization used in medical training helps in delivering highly interactive virtual environment on any personal device. Moreover, this is a less expensive training solution than medical simulators designed for high-fidelity patient simulation which also needs use of real equipment. Compared with simulators, there exist some general disadvantages when it comes to using game engines in training systems (Kuskuntla, Imsand, & Hamilton, 2010). In order to prove the usefulness of games, the authors enhanced and modified all the dynamic effects and existing functionalities that create highly interactive training environments and concluded with effectiveness the use of game engines in medical training systems. A great number of examples using games in medicine can be found in this survey (Wiehagen, 2008). Especially two groups of medical systems, compared with our solution, are interesting: digitally enhanced mannequins and total immersion virtual reality. The advantage of these two types is in using various techniques (e.g., haptics) that combine real and virtual elements. Such medical training systems can be used in combination with X3D (Jung, Recker, Olbrich, & Bockholt, 2008) and are capable to create highly interactive environments with haptic user interface. However, these solutions are very expensive and require the participants' presence in the training process. Our work is based on delivering cost-effective and cross-browser solution that trains medical staff without need for their physical presence in simulated environment.

On the other hand, there is a challenge due to differences between gaming and training preferences (Roman & Brown, 2007). Generally, the main goal of the games is entertainment while for the assessment purpose it can be considered just as a motivational factor. In both cases, the conditions are controlled but the trainee has much more freedom in serious gaming than in assessment based on SGIQ. The other difference is more generic – the competition between trainees in serious games is obvious while in SGIQ-based assessments they act individually and compete with the time limitations, their own misunderstandings or unsatisfactory skills.

Different from serious gaming software which is implemented mainly as a network application, in assessments the highly interactive gaming technology is implemented as a part of a bigger solution. More precisely, the existing assessment system is enhanced with the new type of questions. X3D represents one of the contemporary technologies that enable creation of three-dimensional (3D) virtual reality that can be delivered and executed on commonly accepted Web infrastructure (Brutzman & Daly, 2007). There are several paperworks that describe the ways how this technology can be used for educational purposes. In Mani and Li (2013), regardless of focusing on presenting OpenGL stand-alone solution for surgical training, the authors recognized the X3D capabilities for exchanging the virtual models in standardized way between applications as well as the possibility of integration of these models in interactive Web applications. Much more discussions about the integration of such technologies into Web-based e-learning platform can be found in Di Cerbo, Dodero, and Papaleo (2010). They adapted Moodle LMS (learning management system) to be two-dimensional virtual e-learning space in which the students act as avatars in order to enrich the collaborative e-learning environment with the space model. The main expectation of their improvement was to intensify the collaborative activities among the learners based on obviousness. Represented details of the design and implementation details were found useful for our own solution. Other part of solution is a framework for interpretation of X3D code by the Web platform.

X3DOM formatted content could be interpreted in almost all contemporary browsers due to existence of WebGL plug-ins that can be downloaded for free (Zollo *et al*., 2011). The details about technologies such as Ajax3D that provides intensive data exchange between the Web browser and the server side application, and integration of X3DOM framework in HTML5 DOM format necessary for rendering 3D graphical scenes are well described there. Finally, the results of using these highly interactive technologies are evaluated in Pittarello (2013). Different kinds of scenarios and recurring interaction patterns (e.g., zoom in/out and showing quantitative data associated to 3D objects) were included in the research in which the 50 Computer Science students got them as tasks to implement. Throughout the research, numerous feedbacks are collected and systematized. Client side heavy computational load need for additional functionality and downloading content compression represent the most common ones from the perspective of the study presented in this paper.

Architecture and design for SGIQ assessments

The overall system architecture consists of two parts of almost equal importance. Common client server approach is used due to the choice to develop the Web-based solution. Client side is responsible for graphical rendering rich user–assessment system interaction which includes moving through the virtual space and seeking for correct answers, and finally for intensive data exchange with the server side applications.

Client side basic principles

Web browser with WebGL support represents the very basis of SGIQ client. Fortunately, it is widely supported in the most of contemporary browsers. The SGIQ question is delivered to the client's browser as X3D element (embedded in div element) of the Web page. There are the question scene and related script that provide its specific behaviour. Regular graphical rendering and interaction with the user is provided by the scene set-up and X3DOM functions library (delivered as a separate x3dom.js file). The SGIQ question delivered during assessment is shown in the next example (Figure 1). Its appearance is presented on the left hand side. At first look, it represents the Web form (blue framed panel) commonly used for 'single choice' question design. The question content is below the header part of the form, labelled with question number and followed by the list of alternative answers and

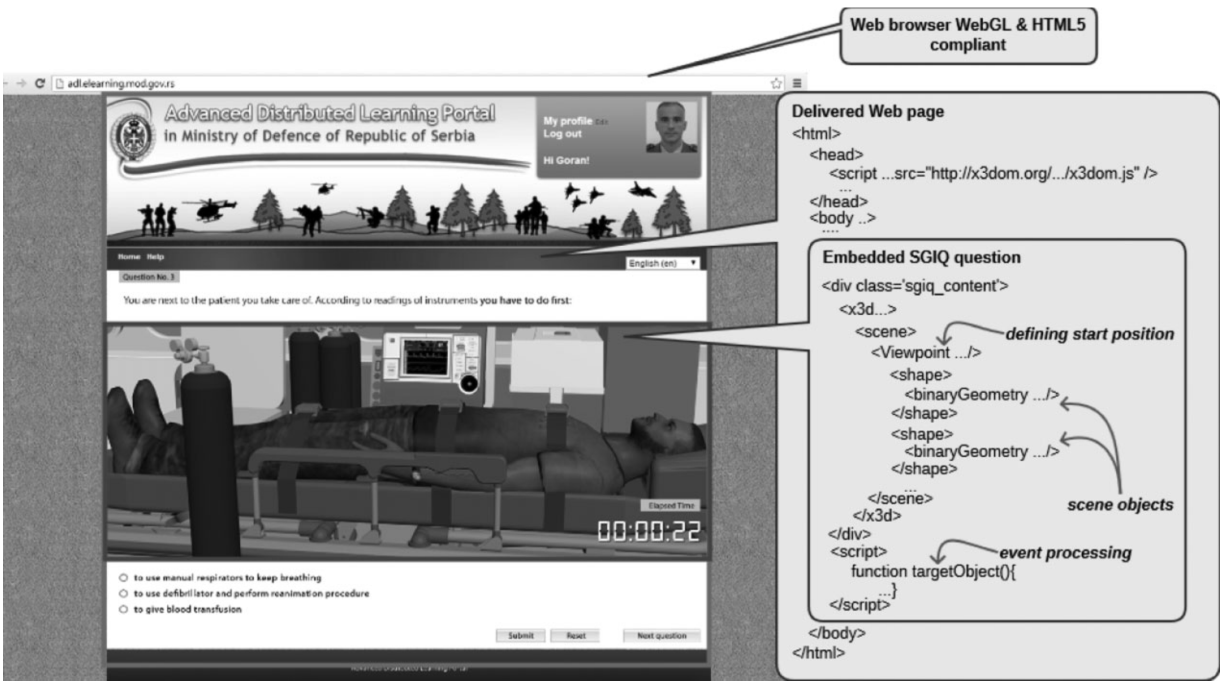


Figure 1 Client Side Interface and Background Script

command buttons. While in the regular 'single choice' question the illustrations, audio or video content, can be used as well as a text, in the SGIQ question the real content is a 3D interactive space while text represents only an explanation of a task to be performed by 'moving' through it for finding information relevant for answering.

On the right hand side are the important technical fragments that reveal the client side logical design. As mentioned, for correct rendering and full functionality, the Web browser should have a built-in WebGL engine. Delivered content exploits this engine in different ways. Delivered SGIQ questions are basically HTML5 documents with the embedded X3D content – the hierarchical structure with the scene element on top which contains child elements for defining objects, their appearances and behaviours. For instance, the purpose of Viewpoint element is to define the trainee's start position (practically 'camera' object). All of the graphical objects are defined as shape elements. In the presented solution, regarding complex elements of scene, these objects are defined as binaryGeometry shapes. This means they are delivered as executable byte codes that directly feed the WebGL engine. In this way, the fastest rendering of scene is achieved and content downloading is optimized. These elements hold the references to the binary files delivered as resources of SGIQ question. Binaries are delivered as separate files while the binaryGeometry elements only contain references to these files, the initial positions and sizes and data necessary for resolving object's transitions. The data exchanging and time controlling, triggered by events such as mouse move, mouse down and mouse up, are implemented separately by the script function targetObject. More detailed consideration can be found in section about client–server interaction.

Server side logic: implemented didactic and client–server interaction

On the server side, there is an assessment system – Web-based application, firstly designed for massive assessments by delivering single/multiple-choice questions to the students (Šimić & Jevremović, 2008). During 7 years of exploitation, the system has been modified several times. Enabling SGIQ-based assessments is the last improvement that has been made. The assessment designed for such specific kind of training is a story-based sequence of questions with the gradual approach. Basically, three types of questions are designed for this purpose. The system dynamically creates them in accordance with the story thread. The scenario is the basic one and the system is designed to add the details sequentially depending on the trainee's behaviour and answers given during the assessment session. The SGIQ-based assessment is designed for evaluating three kinds of skills and knowledge necessary for successfully participating in AMET missions:

- Knowing and fast finding of equipment and medicament in the aircraft.
- Monitoring the patient state during the transport.
- Knowing the procedures and actions appropriate to the patient's state and conditions.

Next are the descriptions of these three question types.

Find Equipment Questions

This is the most original question type due to unique answer design. There are no text fields and areas, checkboxes, radio and submit buttons commonly used for answering the existing question types. There is 3D virtual environment – aircraft interior instead (Figure 2). The answers are represented as objects in this space model. The trainee's task is to find a particular piece of equipment by going through the craft's interior until the 'touch' with the specified object (CoP – Collision of the Positions). The craft entrance always represents the start position in this type of question.

The environment is changeable according to the question content. There are three aircraft models implemented and their names alternatively appear in the question text placed in the position of the first question sign. For AMET missions, all of the crafts have to be equipped with the same standardized equipment set (partially listed in the illustration). As mentioned, the purpose of this question type is to find the particular pieces of this equipment that appears in the text (in a place of the second question sign). The elapsed time is always presented to the trainee due to limited time specified for answering.

The Find Equipment Question (FEQ) type represents the prototype for large-scale independent assessment entities due to the set of the rules attached to

Figure 2 Find Equipment Question Design

every dynamically generated question. These rules enable the question to act as a micro assessment module. The next activity diagram presents how it works (Figure 3). When the trainees start answering, the module starts measuring the elapsed time. Therefore, two simultaneous processes are created. The trainees answer the question by touching the object in the scene (CoP). This event triggers the module to start processing the answer. The correct answer leads the trainee to the next question. If the answer is wrong, the trainee is 'captured' in the same SGIQ module provided with loop-back mechanism built in the module. The same environment is used to evaluate the trainee for other pieces of equipment. This mechanism is used for obtaining the right information about trainees' knowledge, rather to compensate the influence of wrong answer on question result.

During trainees' attempt, the host system (Web browser) measures the elapsed time and the attempt is interrupted if the time expires. Finally, it is obtained by design that trainee may have more than one attempt. In this case, the environment (aircraft) is changed and the new attempt acts as a new question. It is obvious that in one SGIQ question there can be more than one result. All of them are stored as row data (original form) collected during this assessment's micro session: answer correctness for particular equipment and environment, elapsed time, and navigation path.

Monitoring Patient Questions

The purpose of 'monitoring patient' type of questions is to evaluate the trainees' skills in reading the data from instruments for monitoring the patient's conditions. Different from the previous one, in this question type the interactive environment is a part of question, rather than answer (Figure 4). The answer is given in the 'filing the blanks' form. Trainee is briefly introduced to the task. There are several measurements important for critical medical care which the trainee can read from the equipment and the most important ones are shown in illustration.

The focus is not on orientation in this question type. Nevertheless, the interactivity of this question is in moving of trainee through the virtual interior space to be closer to the appropriate instrument. Trainee uses the readings to fulfil the answer. In this way, the monitoring questions are more complex than previous ones due to need that the trainee has to recognize the appropriate equipment and data for collecting requested

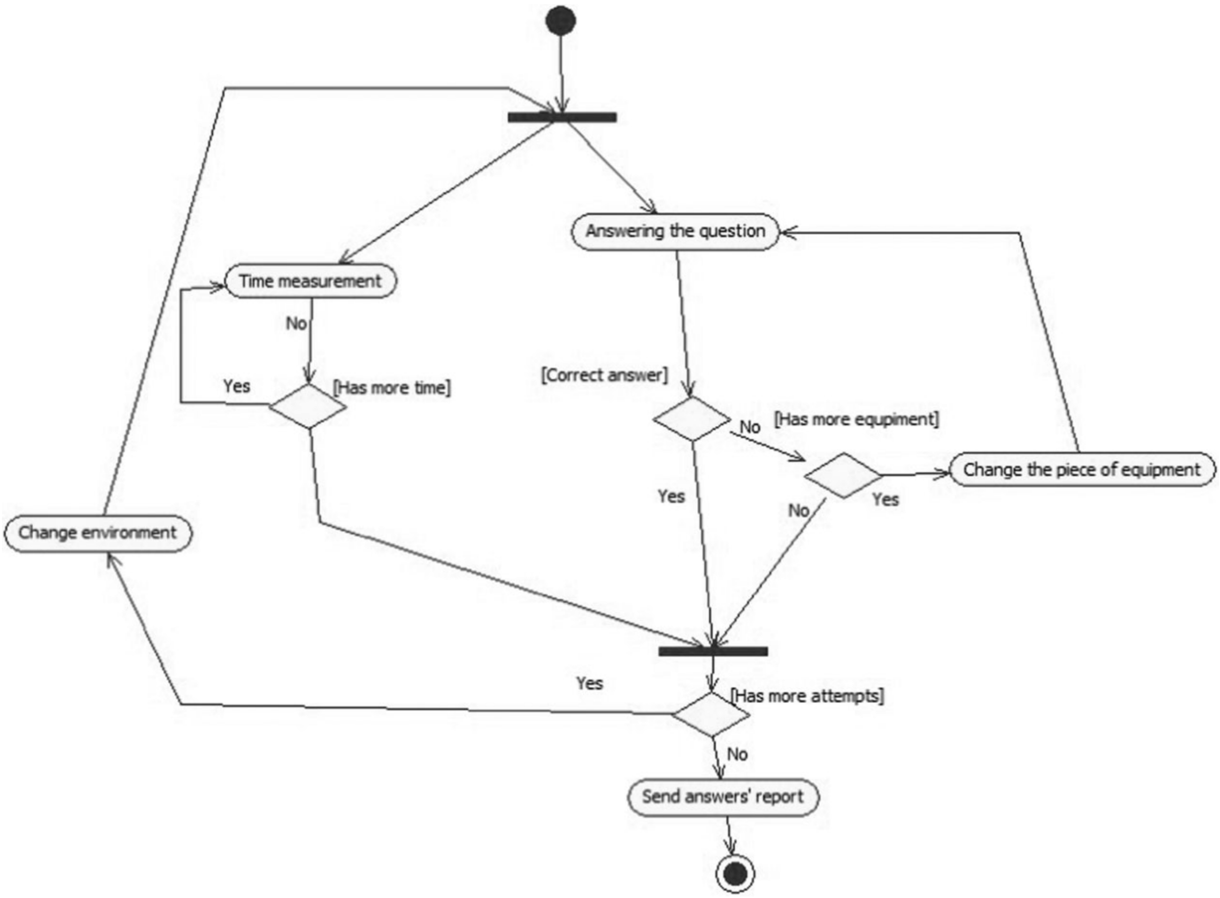


Figure 3 Didactical Principles Implemented in FEQ Tasks

information. The tasks are dynamically generated and answers are checked by simple pattern matching. Monitoring questions can be used in 'repeating' mode in the same way as in the FEQ SGIQ.

Critical Care Questions

'Critical care' question type is designed for evaluating the ability of trainees to be proactive during the AMET missions. The task is to collect all possible information about the patient's state by reading the instruments and other types of equipment and to make decision about which action is the most appropriate at the moment. It is the most complex type due to vast background knowledge implemented for generating the question. There are two variants of this question type – one is with the solutions offered in advance (single choice questions) (Figure 5) and another one is with 'filling the blanks' answer. The first subtype answer is easy to check. More difficult is checking of the free form answer.

The evaluations performed by simple pattern matching brought unexpected results. Therefore, the answers are firstly prepared (pre-processed) for measuring (e.g., elimination of stop words and stemming) and then lexical similarity measure based on word/phrase/sub-sentence comparison is used. More considerations about processing the answers are out of scope of this document. Achieved results are better but still unreliable. Therefore, the single choice variant is in use.

Interaction model

The SGIQ assessment system architecture is developed to support intensive client–server interaction necessary for FEQ tasks. There are two entities on the server side: assessment management server named 'eTest' and logging server named '3Dlogging' (Figure 6). They are

Figure 4 Monitoring Patient Questions Design

dedicated to serving different types of row data obtained from user–system interaction. The interaction begins with the request for SGIQ task, which is immediately delivered to the client (step 1). The events generated in 3D environment that are mentioned above are recorded on '3Dlogging' server (step 2). The client collects the next event parameters: session id, position x/y/z, orientation x/y/z, event type, current time stamp and additional data (optional). Session id is used for associating with the particular assessment session. Approximately 150 bytes of data are sent from the client side for each event without using optional data, providing the effective way for using a link capacity.

Events data are buffered and periodically sent in form of JSON-formatted messages by using AJAX calls. The reason for buffering is that the need for synchronization due to trainees' movement through 3D environment produces a large number of events that cannot be processed immediately. In practice, it would be ineffective or even impossible to create that frequency of AJAX calls. The 'mouse button up' events are used as a trigger for sending the buffered data because the movement through space is performed by moving a mouse with the mouse button pressed. The data recorded in this way are sufficient to completely reconstruct the 3D session from beginning to end. Mouse 'click' event is used for sending the answer. It is performed by a separate AJAX call directed to 'eTest' server (interaction No. 3).

In this way, the data are processed through the two channels. In the back end of the server side, they are aggregated again for providing the final trainees' results to the teachers. Additionally, more detailed analysis could be performed on 3D sessions data in order to get appropriate information about students' behaviours and to make conclusions on this.

Figure 5 Critical Care Questions Design

After assessment analysis

The presented assessments have been conducting with the group of six to eight trainees (usually two or three physicians and four or five medical technicians) two times a year as a part of the final preparations for AMET missions. Since SGIQ assessments have been used since the spring 2013, the results of the two groups are presented (a total of 14 trainees). One of the demands was to evaluate the usability of this approach

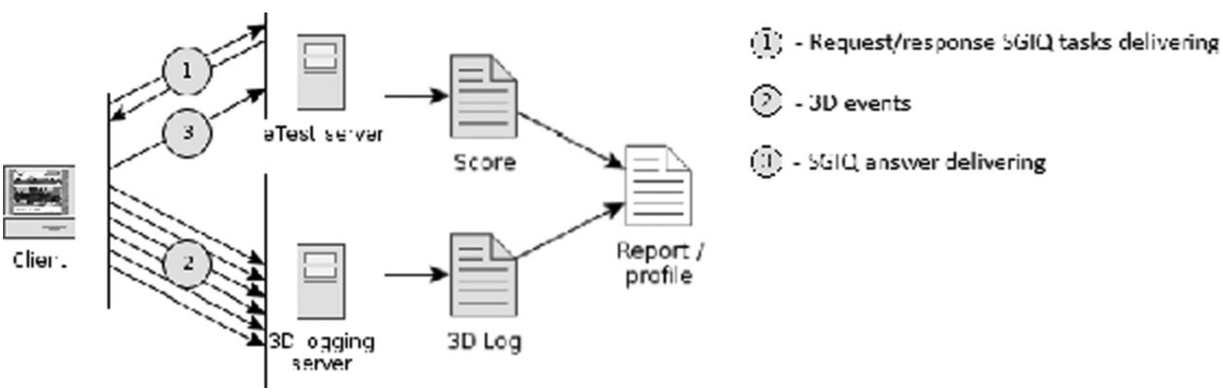


Figure 6 Interaction Model

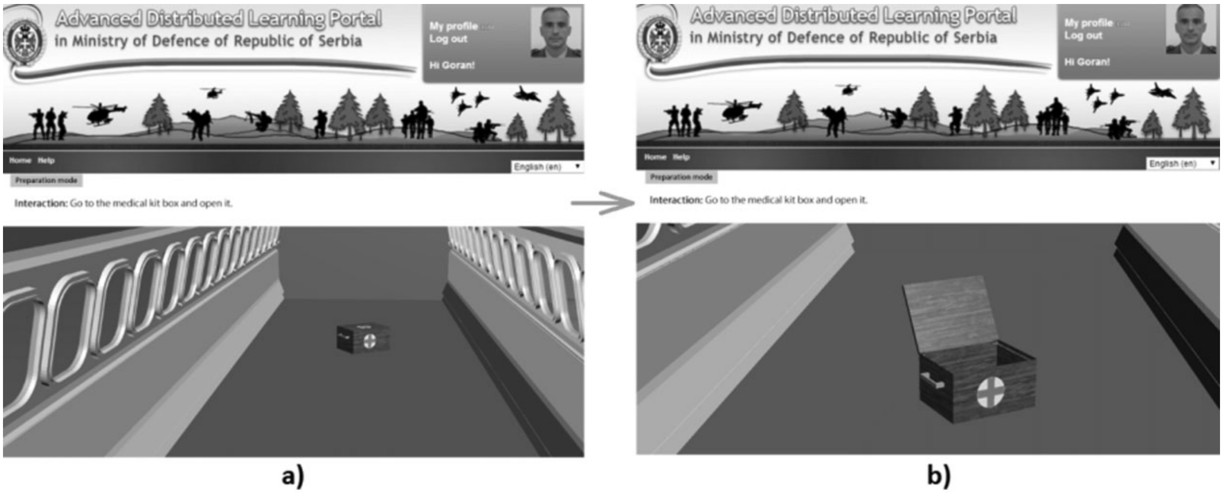


Figure 7 Preparing Task: a) Scene at the beginning and b) Scene at the end of the task

in terms of efficiency, effectiveness and satisfaction (according to ISO 9241 part 11). Generally, these three terms are considered as: efficiency represents the amount of resources expended during the task completion; effectiveness is the number of successfully completed tasks; and satisfaction represents a comfort and acceptability of the work system to its users and other people affected by its use. For measuring purposes, they are used as follows: for measuring efficiency, two main dimensions are used – time and length of path taken to complete a task; measure for effectiveness is a trainee's score represented as ratio between opened and successfully completed tasks; finally, for measuring satisfaction, we used survey on trainees after finishing the test.

Efficiency and effectiveness

As we have defined efficiency as time and length of path taken to complete a task, in this section we present the results of the analysis of 14 trainees' behaviours within the assignment with one preparing and 15 regular tasks. During alpha tests (Van Veenendaal, 2012), it was recognized that candidates should be prepared for using the new way of assessments. Effectiveness measured in a way described above was less than 51% in assessment as a whole. Through detailed analysis of results, the common characteristic was discovered, that the first task delivered was the worst one (success was less than 20%) and the trainees have been more successful in every next delivered task. It was recognized that the trainees had a problem with movement through the space model used in tasks. By introducing the separate, so-called preparing tasks – the simple ones whose goal is to reach 10 m distant object and interact with it within 10 s (opening the box) brought the significant improvement of efficiency (Figure 7). Solving this task successfully made trainees familiar with navigation in virtual 3D environment and made them prepared for real assessment. All 14 candidates solved this task within three attempts (eight of them did it in first attempt and four of them in second attempt). The average time and path taken for this task (analysing successful attempts only) are 9.16 s and 10.87 m.

After the preparing task is fulfilled, the trainees can start assessment. Fifteen tasks (five of each type) were delivered to each trainee. In this way, the first five delivered questions were of FEQ type, followed with five MPQ (Monitoring Patient Questions) type tasks and five CCQ (Critical Care Questions) type tasks at the end. In performing FEQ-type tasks, the results of all 14 trainees were successful due to use of preparation tasks and appropriate knowledge level (it was easy for experts to recognize the pieces of equipment). Therefore, the only relevant information for evaluating their efficiency for this type are navigation paths and time consumed (better result for shorter time and path is obtained). In the next two diagrams (Figure 8), the full number of 210 considered results is shown. In the left side diagram (a), path and time row data are put in correlation. Each trainee has 15 'answers' (results) and

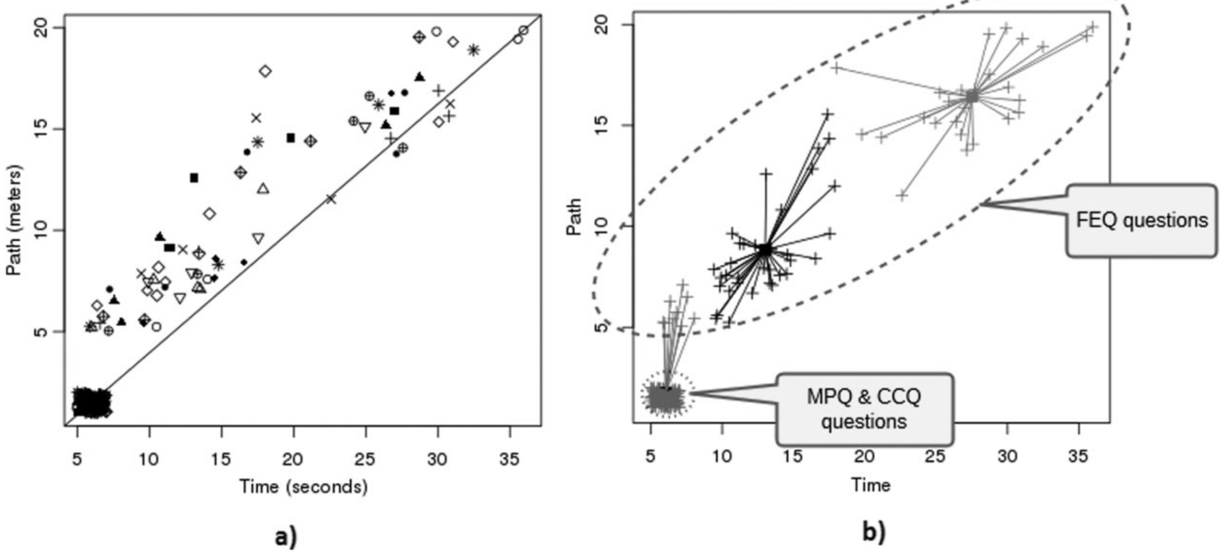


Figure 8 Efficiency Analysis by Using Correlation and Clustering of Data: a) Correlation of data and b) Clustering of data

they are represented by a separate symbol. Strong correlation between time and path length is obvious and expected. On the other hand, the most important information is in distribution of results. It is not a uniform one and we tried to find its patterns. For that reason, the K-means clustering (MacQueen, 1967) of results was performed by presumption that it should be three clusters depending on three types of questions. The R framework is used for this purpose (Kabacoff, 2011). This process brought the results presented in the right side diagram (Figure 8b).

Formed clusters are shown by their centroids and the results dots were represented as crosses connected to the centroid of cluster to which it belongs to (green, black and red). After the detailed analysis of results, it is discovered that the results for the MPQ and CCQ types of questions are strongly grouped close to the origin of coordinate system (dotted ellipse at the bottom left corner). It means that time and paths are not relevant for considerations about the results of these question types. The trainees had excellent results on MPQ and CCQ – 93.5% of the correct answers (only 9/140 wrong answers). Two reasons are found for this success – previous experience and high level of domain expertise of trainees is the first one; the achieved effectiveness is a result of final assessment is another one. If SGIQ had been used during the Learning and Training (L&T) process the results would probably have been worse.

In contrast, time and paths represent the relevant information for FEQ-type results. The FEQ results (dashed ellipse) belong to all of the three clusters. Regardless of such distribution seems appropriate for evaluating the trainees' efficiency in FEQ task, deeper analysis is needed due to differences in aircrafts used as task environment. As mentioned, during the assessment there are five FEQ tasks delivered to each trainee – two in helicopters and three in airplanes. The distribution of results regarding these two subtypes (Figure 9) points out the significant differences although the paths are double longer in airplanes (crosses), the times spent for finding the piece of equipment in airplane are in many cases three to four times longer than those in helicopters (triangles). Instead of linear, the dependence

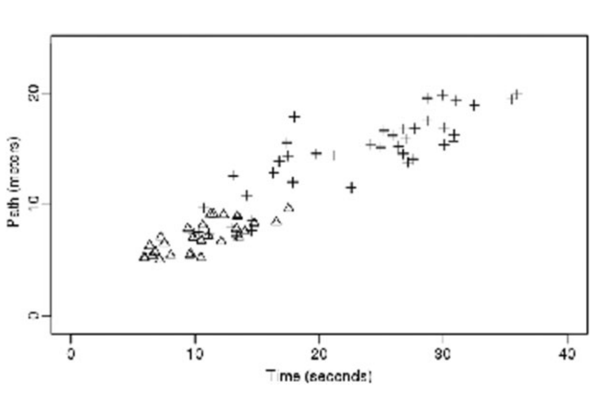
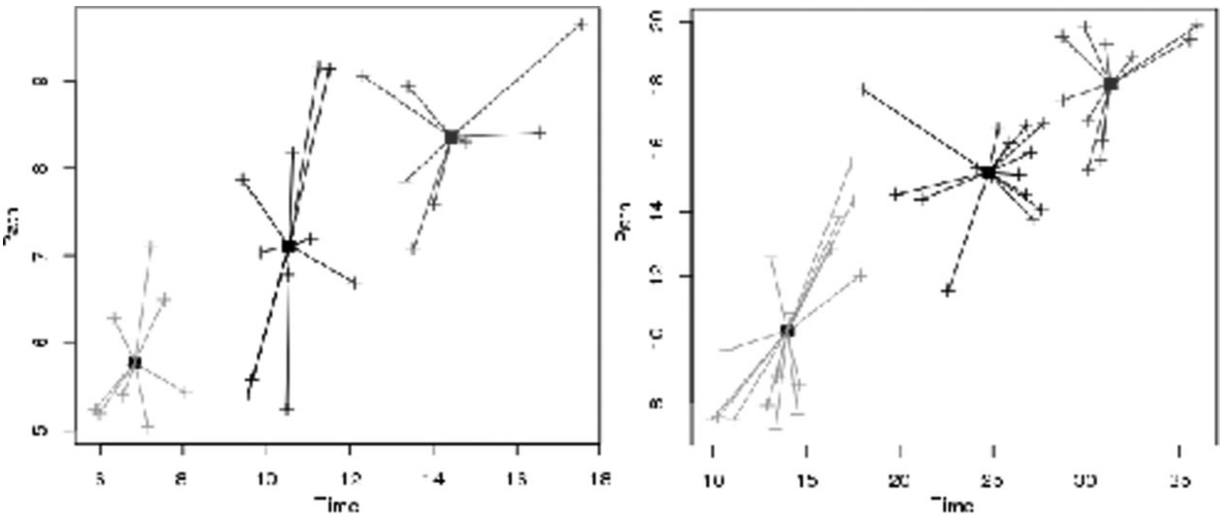


Figure 9 Different Efficiency in Helicopters and Airplanes



a) b)

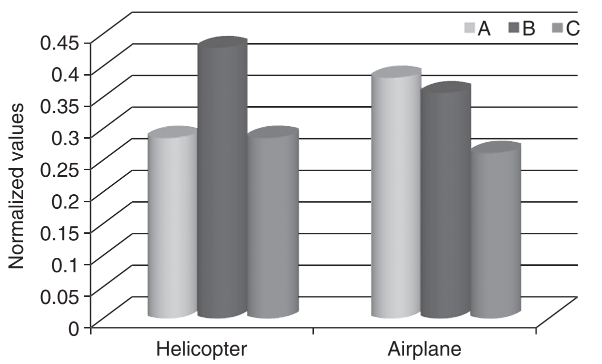
Figure 10 Clustered FEQ Tasks Results for (a) Helicopters and (b) Airplanes

between these two dimensions (path and time) is almost exponential. Further, it was discovered by analysing the particular sessions that the seats in the airplanes (because regular airliners, modified by lowering the seats, are dominantly used for AMET) represent another difficulty which additionally reduces the efficiency of trainees. As in the real conditions, they are causing the collision during a movement through the rows. Some of high-skilled trainees obtained the extraordinary results in airplanes almost better than the helicopter ones. Nevertheless, most of them are worse especially in time differences.

Because such representation has no sufficient clarity, the FEQ results were analysed separately. Depending on subtypes, they were split in two groups (FEQ in helicopters and airplanes). Further, they are clustered in the same way as mentioned above. The clustering results are represented in the next two diagrams (Figure 10).

The similarity between results obtained in helicopters (a) and airplanes (b) is obvious. The best results are grouped in the bottom left corner (green dots) while the worst ones are in the upper right corner (red dots). In the middle are the rest of them (black dots). For this reason, three marks are formed: ‘A’ – the best one, weighted by 3; ‘B’ – middle one, weighted by 2; and ‘C’ – the worst one, weighted by 1; they are used for evaluation and analysing the results. The distributions within the subtypes of FEQ tasks are shown in the next diagram (Figure 11).

Fourteen trainees made 28 helicopter (two tasks) and 42 airplane (three tasks) results. The average results are ‘B’ in both cases and there is minor difference between them (exactly 5.62%). On the other side, the distribution of ‘A’ and ‘B’ results is significantly different in these two cases. There is ‘A’ results difference of 25% in favour of the airplane tasks and ‘B’ results difference of 20% in favour of the helicopter tasks. Because of tasks complexity, the opposite distribution is expected. Relevant for its interpretation is the fact that the tasks are delivered in order – helicopter ones firstly and airplane ones after that. The ‘C’ results difference is minor



	A	B	C	num res
Helicopter	8	12	8	28
Airplane	16	15	11	42

Figure 11 Comparison of the Results in Helicopters and Airplanes

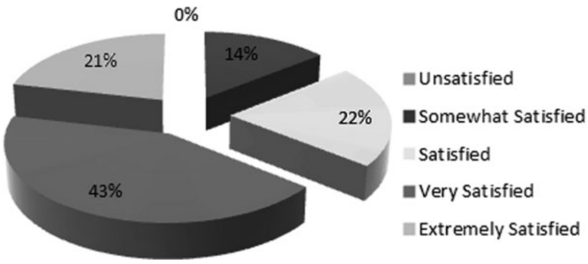


Figure 12 Impressions about SGIQ Assessment

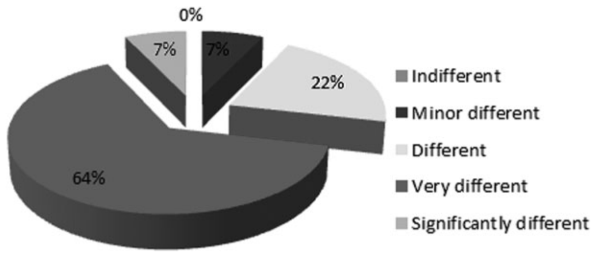


Figure 13 Evaluating Difference between SGIQ and Previously Experienced Assessments

– less than 10% in favour of helicopter tasks, and regarding tasks ordering it has a decreasing trend in favour of other two marks.

Satisfaction

In order to evaluate trainees' satisfaction after the assessment, the survey is conducted. Survey had five questions about different aspects of SGIQ. The first question was to describe the trainees' impressions after using SGIQ by Likert 5-point scale. Values on the scale ranged from 'unsatisfied' to 'extremely satisfied'. Most of the trainees (12) described their experiences by using SGIQ-based assessment as satisfying (86%). No one is unsatisfied while only two of them are partially satisfied (Figure 12).

The second question in survey was to rank the types of SGIQ questions in order of their usability and importance (starting with 1 for favorite type). Regarding the level of subject expertise of the trainees, this ranking is found important for evaluating the SGIQ question types. The first rank for FEQ type was expected but this rank belonged to CCQ type (Table 1). Therefore, additional considerations about ranking were conducted with trainees. In their opinion, FEQ types are impressive at first look due to a lot of interaction needed to find the answer. On the other side, the equipment is usually grouped and easy to find. From their perspectives, CCQ types are more serious because of solving-problem approach and much more domain knowledge implemented.

One of the questions in survey is about subjective impression about difference between SGIQ-based assessments and the others experienced and related to the same domain (Figure 13). The Likert-scaled questions were used for this purpose. Trainees could select options from 'indifferent' to 'significantly different' on five options scale. Results obtained by comparative question are more better than expected and there are 93% positive feedbacks and without negative ones. Just one trainee has opinion that there is a minor difference.

The final question is the most direct – whether the trainee prefers the SGIQ-based assessments or not. Only one of them (7%) prefers assessments without SGIQ-based tasks. For better profiling of the examinees, there was one 'control' question – how many attempts the trainee spent for preparation task. The trainees declared that eight of them had only one attempt, while the rest (six) had more. The answers were put in correlation with the ones of the other questions.

Conclusions

Assessments based on SGIQ are likely to become the regular ones used for training of individuals for critical medical care during the transport and/or evacuation of critically ill patients. Different from serious gaming environment, in which everything is interactive and visualized, in SGIQ assessment the virtual environment is partially used in accordance with strictly

Table 1. Ranking of Questions Based on Trainees' Preferences

Ranks				
1	2	3	Result	Questions
6	7	1	33	AE&E (Alternative Environment & Equipment) Questions
0	1	13	15	Monitoring Patient Questions
8	6	0	36	Critical Care Questions

didactic purpose and rules. The presented case study is focused on the principles implemented in the new question types. The question content is dynamically created for each type. Nevertheless, the domain ontology developed is out of scope of this document regardless of considerable efforts made. For instance, recommended transport kit comprises 67 pieces of equipment and minimum of 44 necessary medications – no one is redundant and they all have different purposes. Although there are many types of illness and injury that the transported patients could have, there are few indications that are important to deal with during the transport: a blood pressure, percentage of oxygen, and heart dynamic, and performances are the most important. Knowing the equipment, their position in the vehicle, monitoring the patient condition and reacting if necessary are required for medical personnel to master them. Therefore, these skills are in focus of SGIQ assessment.

SGIQ-based assessments have brought several benefits regarding the previous ones. Without SGIQ assessments, examining the level of skills mentioned above requires real equipment and use of simulator which generates signals to be measured. Both the equipment and the simulator are expensive as well as their maintenance. Moreover, the equipment is usually used for performing real medical care. For this reason, there are often very few of these assets that implicate the number of trainees who can be tested simultaneously.

Regardless of the fact that X3D is an ISO standard, we were concerned about a fact that the little progress has been made in the last few years concerning the improving of documentation and supporting tools. Developing complex scene parts (e.g., medical care equipment) by simply arranging the X3DOM elements (tags or nodes, as they are called in X3D) in scene graph was extremely time consuming. Although there have been attempts in creating X3D content-editing software (e.g., BS Content Studio), the scenes used in SGIQ tasks were visualized by different tools (3D modeling and animation software such as 3dsMax, Maya and Blender). Concerning the fact that the X3DOM framework with all exporting possibilities has been still under development, there were lots of 3D modeling features that could not be properly represented (exported). For this reason, additional X3DOM scene corrections were done. Experience with creating and exporting airplane interior suggests that polygonal mash structures must be done only in aforementioned software (collaborative parts creation in different tools and assembling scene in one tool result either in no export or in destroyed model structure). Use of optimizer which applies advanced compression techniques resulted in exported aircraft interior becoming dynamic X3DOM scene with minimal structure losses.

The evaluation of the system is performed by following ISO 9241, part 11 recommendations, focusing on efficiency and satisfaction. Despite the difficulties in using X3D technologies, very positive results of survey conducted on 14 trainees (satisfaction) encouraged us to continue with developing SGIQ-based assessments. Apparently, useful information is obtained by ranking question: although it is expected that the best rank should belong to FEQ-type tasks due to their dynamic and highly interactive nature of finding pieces of equipment in the limited time and space conditions, the best ranked questions are of CCQ type. Efficiency is evaluated in more complex way: obtained results are firstly put in correlation, then clustered by using K-means algorithm, and finally the clustered data are analysed for making conclusions. By clustering, it is proved that the time and path are only relevant for FEQ-type tasks. Therefore, further considerations are a concern on them. Firstly, the FEQ results are split into two subsets depending on the environment, and then they are clustered in the same way again. Because both result subsets have a similar distribution, the similar clusters are obtained. After grouping in clusters and normalization, they are compared again. Regardless of the results that are insignificantly different, the distributions of ‘A’ and ‘B’ results in helicopter and airplane subsets are different (20%–25%). Because the order of questions could cause it, verifying such distribution in the other conditions (i.e., better question shuffling in the future assessments) represents one of the next research tasks.

The trainees suggested that it would be useful if the SGIQ tasks were included in the other parts of the training process. Therefore, future research will include their integration in the existing learning management system. Although SGIQ supports the cognitive aspects of assessments, there are several non-cognitive concepts such as emotions, feelings, perceptions, motivations and interests that affect the results. One of the efforts in the future research will be developing ways to identify and measure them, recognize the patterns of behaviour, and integrate them into considerations about results. Finally, for the next generation system, it has been planned to incorporate more interactivity in dealing with the patient and to include more audiovisual effects in scenarios according to the ones that our medical teams already experienced (e.g., noisy engine sounds, air turbulence, night fly conditions, aircraft defects). A more realistic scenario produces the better learning effects. Unfortunately, it is not possible to reproduce the real-world situations with all their details. Environment and equipment designers represent the critical resource for such improvements. For instance, the visualization of a trembling patient who starts bleeding, changing the expression and trying to say something, needs several dozens of hours. One of the main future goals will be creation of complex patient model which could be reusable for different scenarios providing different behaviours and appearances. Such investment is expecting to provide production of SGIQ in a much effective way.

References

Brutzman, D., & Daly, L. (2007). *X3D: Extensible 3D graphics for Web authors*. San Francisco, USA: Elsevier Morgan Kaufmann.

Di Cerbo, F., Dodero, G., & Papaleo, L. (2010). Integrating a Web3D Interface into an E-learning Platform. *Proceedings of the 15th International Conference on 3D Web Technology, Association for Computing Machinery*, pp. 83–91.

Jung, Y., Recker, R., Olbrich, M., & Bockholt, U. (2008). Using X3D for medical training simulations. *Proceedings of the 13th international symposium on 3D web technology (Web3D '08)*, pp. 43–51.

Kabacoff, R. (2011). *R in Action – Data analysis and graphics with R, manning*. (pp. 263–290). Shelter Island, NY: Manning Publications Co.

Kuskuntla, R., Imsand, S. E., & Hamilton, J. A. Jr. (2010). Enhanced expert field medical training simulations and their effect on the modern combat life saver training procedures. *Proceedings of the 2010 Spring Simulation Multiconference (SpringSim '10), Society for Computer Simulation International*, Article 49, 6 pages, USA.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pp. 281–297, University of California Press, Berkeley, California.

Mani, G., & Li, W. (2013). 3D Web based surgical training through comparative analysis. *Proceedings of the 18th International Conference on 3D Web Technology*, ACM, pp. 83–86.

Pittarello, F. (2013). Testing the X3DOM framework for the development of Web3D applications. *Proceedings of the 18th International Conference on 3D Web Technology, Association for Computing Machinery*, pp. 191–194.

Roman, P., & Brown, D. (2007). Constructive simulation versus serious games: A Canadian case study. *Proceedings of the 2007 Spring Simulation Multiconference (SpringSim '07)*, Volume 3, pp. 217–224.

Šimić, G. (2012). Constructive simulation as a collaborative learning tool in education and training of crisis staff. *Interdisciplinary Journal of Information, Knowledge, and Management*, *7*, 221–236.

Šimić, G., & Jevremović, A. (2008). *E-learning tools embedded in the University information system*. Retrieved from http://www.researchgate.net/profile/Goran_Simic2/publication/266891825_E-learning_Tools_Embedded_in_the_University_Information_System/links/5451de3c0cf2bf864cbaa3c9.pdf

Van Veenendaal, E. (2012). Standard glossary of terms used in Software Testing. International Software Testing Qualifications Board, pp. 8–9.

Warren, J., Fromm, R., Orr, R., Rotello, L., & Horst, M. (2004). Guidelines for the inter- and intrahospital transport of critically ill patients. *Critical Care Medicine*, *32*(1), 256–262.

Wiehagen, G. (2008). Medical Simulation and Training – Breakthroughs and Barriers. *US Army Medical Research and Materiel Command, Telemedicine and Advanced Technology Research Center*.

Zollo, F., Caprini, L., Gervasi, O., & Costantini, A. (2011). Web3D '11. *Proceedings of the 16th International Conference on 3D Web Technology*, pp. 129–136.

Zyda, M. (2005). From visual simulation to virtual reality to games. *IEEE Computer*, *38*(9), 25–32.