Journal of Universal Computer Science, vol. 23, no. 2 (2017), 208-232
submitted: 15/3/16, accepted: 2/2/17, appeared: 28/2/17 © J.UCS

A Cost-Effective Approach for Procedural Training in Virtual Worlds

Mariano Rico
(Department of Artificial Intelligence, UPM, Madrid, Spain
mariano.rico@upm.es)

Jaime Ramírez
(School of Computer Engineering, UPM, Madrid, Spain
jramirez@fi.upm.es)

Diego Riofrío-Luzcando
(School of Computer Engineering, UPM, Madrid, Spain
driofrio@fi.upm.es)

Marta Berrocal-Lobo
(CBGP, INIA, School of Forestry Engineering and Natural Resources, UPM
Madrid, Spain
m.berrocal@upm.es)

Abstract: Virtual worlds are one of the most promising and interesting lines of research and development in e-learning. Also, the recent maturity of 3D virtual world platforms like Second Life or OpenSimulator has made it possible to deploy such systems at a low cost in schools, colleges or companies. Therefore, these virtual world platforms have enabled the development of cost-effective virtual laboratories. However, there are few empirical studies to support the educational value of Virtual Worlds for Procedural Training (VWPT), especially when they are equipped with automatic tutors. In this paper we present an empirical study that provides evidence in favor of the educational value of virtual worlds. To be precise, we describe the evaluation of a biotechnology VWPT implemented on top of OpenSimulator. We have implemented an automatic tutor within this virtual laboratory in order to supervise and guide students during task practice. This tutor is configurable and can be reused in different VWPTs. This facilitates the development of VWPTs. This virtual laboratory has been evaluated by domain experts (university professors) and students over a three-year period as part of a university course. The analysis of this evaluation has shown good results for usability, user satisfaction and perceived educational utility. It has also served to demonstrate that the virtual laboratory is resilient and reliable. This supports the applied technical approach.
Key Words: virtual realities, virtual worlds, OpenSimulator, procedural training, computer-assisted instruction, biology and genetics.
Category: J.0, K.3.1, I.3.7, L.2.3, L.2.7, L.3.0 L.3.6, L.5.0

1 Introduction

A virtual world is a particular instance of a 3D virtual environment. It can be visited via the Internet by several users at the same time using avatars (their

alter egos in the virtual world) or a first person view. Typically users can navigate around the world, interact with objects (touching, taking, pushing items, etc.) or chat with others users of the virtual world.

Virtual worlds are signified as being one of the most promising and interesting lines of research and development in e-learning [Dalgarno 2002, Messienger et al. 2009, Duncan et al. 2012]. Virtual worlds can be used in different educational applications. One of them is to train students in procedures, for example, they can support task practices on magnetic fields and magnetism [İnce et al. 2015] or teach children about pedestrian safety [Gu et al. 2015]. We will refer to these specific virtual worlds as Virtual Worlds for Procedural Training (VWPT).

In this paper, we present an empirical study that provides evidence in favor of the use of VWPTs. To be precise, we describe the evaluation of a biotechnology virtual laboratory implemented on top of OpenSimulator. This virtual laboratory is equipped with an automatic tutor that can supervise and guide students during task practice. This tutor is configurable and can be reused in different VWPTs. This facilitates the development of VWPTs.

This virtual laboratory was evaluated by domain experts and students. The results of the study went beyond the performance of the automatic tutor, because they led to some findings of interest on the development and use of virtual worlds. As part of this evaluation, experts and students filled in a questionnaire after task practice in the virtual laboratory.

The structure of the remainder of the paper is as follows. Section 2 discusses relevant works and related approaches in this field. Section 3 describes the biotechnology virtual laboratory. In Section 4, we describe the main features of the automatic tutor that is integrated with the virtual laboratory. Section 5 explains some issues related to the development process of the virtual laboratory. We then outline an evaluation of the virtual laboratory as a whole (i.e. including the automatic tutor) through an experiment with biotechnology experts (Section 6) and an experiment with students (Section 7). These two sections report and discuss the results of the experiments. Section 8 then presents a critical analysis of the technical and pedagogical approaches followed in this paper. Finally, we outline the conclusions of this research in Section 9, and we end with some future lines in Section 10.

2 Related Work

Nowadays a wide variety of educational virtual environments are readily available. Some use pictures, videos or interactive objects to display information for the students or help teachers to give virtual lectures. Of the learning activities defined by Dalgarno [Dalgarno and Lee 2012], we will focus on task practice because the presented virtual laboratory supports this activity. We will confine our

survey to the training in closed-ended tasks, where there is only one procedure to achieve a valid final state or more than one similar procedure with minor differences between each other.

The use of virtual worlds to acquire procedural knowledge is the result of a constructivist view of learning [Dalgarno 2002], according to which learning occurs when, during the active exploration of the knowledge domain, students discover a gap in their knowledge or an inconsistency between their knowledge representation and their experience.

Within the VWPT group, we will make a distinction between applications that are or are not equipped with an automatic tutor. Firstly, we will present some applications that do not have an automatic tutor but have reported user evaluations. One of the most well-known of these applications is a NASA VE to train astronauts in extravehicular activities (repair and maintenance missions) on the Hubble Space Telescope [Loftin and Kenney 1995]. Another remarkable system presented in [Brown et al. 2012] trains nurses for the Intensive Care Unit, and another called SWIFT (Second World Immersive Future Teaching) [Rudman et al. 2010] implements a virtual genetics laboratory in Second Life. Only a few works in this group have been tested in an university setting. Of these, we can highlight the Istanbul University Virtual Laboratory (IUVIRLAB) [İnce et al. 2014, İnce et al. 2015] and iVirtualWorld [Zhong and Liu 2012]. iVirtualWorld is more than just a virtual laboratory, since professors can use the tool to create their own chemical experiments as a virtual laboratory.

On the other hand, VWPTs equipped with automatic tutors can typically supervise student performance, detect their errors and provide hints on demand. One of the best-known VWPTs of this type is *STEVE* [Johnson and Rickel 1997], a 3D animated agent that highlighted the advantages of using a 3D tutor avatar for procedural training and was later utilized in a project called MRE (Mission Rehearsal Exercise) [Hill et al. 2006]. Other examples of such VWPT systems are MASCARET (firefighter operational management) [Buche et al. 2003] and SafeChild (child pedestrian safety) [Gu et al. 2015].

No empirical evaluations with users have been reported for any of the above VWPTs with automatic tutors. This suggests that there is a significant gap in this field, i.e. a shortage of empirical studies to support the educational value of VWPTs. SafeChild's authors reported an empirical study, but they did not employ the tutor in that study.

One exception to this rule is a well-known project called Lahystotrain [Los Arcos et al. 2000], developed to train surgeons in laparoscopy and hysteroscopy operations. It was validated by experts and students by means of a questionnaire. Another exception is TRANSoM [Pioch et al. 1997], which is a virtual environment (VE) for training pilots of remotely operated underwater vehicles (ROV). It was evaluated [Fletcher. 1998] by means of a training transfer experi-

ment. This experiment analyzed the learning effectiveness of the VE for training in maneuvering skills by means of pre- and post-tests. These tests were administered using a ROV.

As we will discuss later, the main contribution of this work is not the design of the automatic tutor, which has many features in common with previous systems, but the evaluation of the virtual laboratory with experts and users. We think that this evaluation supports the educational value of VWPTs in a field where evidence is scant.

Given that our virtual laboratory has been implemented on OpenSimulator, it is also an example of how to develop cost-effective educational applications that can be deployed at schools and universities.

3 A Biotechnology Virtual Laboratory

The aim of the task practice created for this laboratory is to enable students to study the function of a gene encoding a protein responsible for protecting poplar trees against diseases. To do this, each student has to use his/her avatar to rapidly produce and genetically transform a large number of poplar plant progenies in the virtual laboratory to study the function of this protein. The virtual laboratory consists of a building divided into five rooms: hall, main room, autoclaves room, phytotron and laminar flow cabinets room. Students perform much of practice in the main room where most of the chemicals and much of the equipment is available. Some examples of the equipment required for the practice are beakers, tubes, ice trays, ovens, micropipettes, etc. Typically, undergraduate students do not have access to biotechnology laboratories with this kind of material. In any case, this task practice would have a cost of from 5,000 to 25,000 euros per student in real life (these data were provided by the professor who designed the practice). This is unaffordable for a higher education institution. Moreover, the whole practice would typically take two years in real life. This is much longer than the length of a course. This virtual laboratory overcomes these two problems by speeding up the experiment as much as necessary and by removing any cost derived from the use of expensive laboratory material. Videos of experiment participants using this laboratory are available for viewing at `http://3dlabs.upm.es/laboratorios.php#lab16` and `http://youtu.be/mAFREZ5_iak`.

The task practice is designed as a closed-ended problem in the sense of that there are not many ways to achieve the final goal. Students have to follow a pre-established protocol, although, at some stages of the practice, they are free to take actions in different orders, for example, when they have to take chemicals from different places.

The learning activity combines highly and lowly structured parts, carefully selected by the professor who designed the practice. In the highly structured

parts, the student is told which action he/she should take next, and he/she is only allowed to make irrelevant errors. On the other hand, in the lowly structured parts of the practice, the student is given a only general hint instead of being told specifically which action to take next. For example, when the student has to prepare a mixture of chemical components, he/she is told to prepare a mixture, but not which specific chemicals are required. This learning-design decision is consistent with Productive Failure [Kapur 2008], and we agree that it results in improved learning.

Our biotechnology virtual laboratory borrows some features from computer games: the practice is divided into phases/challenges, that is, the student must complete each phase/challenge before he/she can move on to the next phase. At the end of each phase, the student is informed of his/her mistakes, if any. For pedagogical reasons, however, the system may allow students to pass on to the next phase of the practice even if they make mistakes. The activity of the students in the virtual laboratory is logged for later examination by the professor or the development team.

4 Description of the Automatic Tutor

Our automatic tutor validates the user attempts at actions and provides help while they are performing a procedural task. However, the automatic tutor is not embodied as an avatar in the virtual world.

The user is assisted by means of messages indicating the next actions to be taken in the practice (see top left snapshot in Figure 1). Big balloons are zoomed images of the error messages warning students about wrong actions or objects required in the next action of the practice. However, these messages are also used to show reports at the end of each phase outlining any errors that the student has made so far (bottom snapshot in Figure 1).

Exactly what type of assistance the student will be given at any time is configured at design time in a "tutor configuration" file whose content is easily understandable for a teacher without programming skills. The syntax of this file is described in Section 4.3. As this kind of assistance can be configured, it is possible to adjust the difficulty level to adapt the practice for students with different backgrounds. This configuration file describes the practice protocol in terms of dependencies between student actions. Note that in biotechnology jargon, both the practice procedure and the practice guidebook are named 'protocol'. Figure 3 shows a diagram of part of the practice protocol used in this specific virtual laboratory. Figure 2 details the appearance of the virtual objects involved in the actions illustrated in Figure 3. In this part of the practice, the student has to prepare a mix of three chemicals in a tube. These chemicals are kept in a refrigerator. Firstly, the student has to take an ice tray (action a). Then he/she has

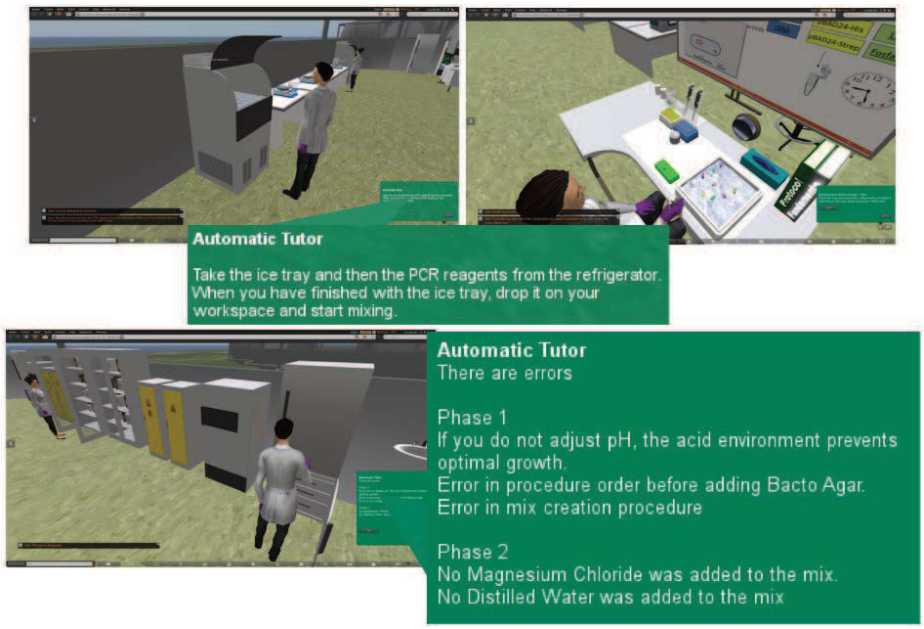


Figure 1: Examples of tutoring messages in the virtual laboratory. Big balloons are zoomed images of the snapshots.

to take these chemicals from the refrigerator (actions b, c and d) and carry them in the ice tray to the workspace (action e) where the mixture will be prepared. Next, the student needs to use a micropipette with a tip (actions f and g) to add each chemical to the mix (actions h.a, h.b and i). Finally, the student takes and carries the tube with the mix (action j) to another part of the laboratory. As can be deduced from the dependencies outlined in Figure 3, there is more than one possible right sequence. Figure 3 shows two possible right sequences, but there are, in fact, 12 possible right sequences. For pedagogical purposes, the system must also allow the student to perform wrong actions such as adding chemical 4 to the mix (action y).

4.1 Action dependencies

Different kinds of dependencies can be specified in the practice protocol:

4.1.1 Regular dependency

A regular dependency expresses that the performance of one action is a requirement for the performance of another. In Figure 3, for example, if we want to

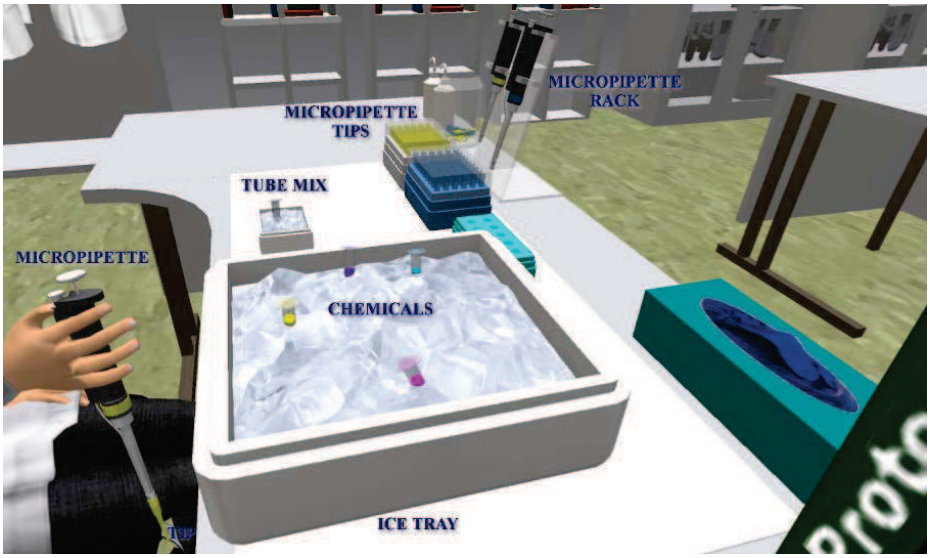


Figure 2: Screenshot of materials used in the example illustrated in Figure 3.

take chemical 1 (action b), we first need to take an ice tray (action a).

4.1.2 Mutex actions

There is another kind of dependency that represents actions that cancel each other out (mutex actions). For example, "switch on a light" and "switch off a light" or "take ice tray" (action a) and "drop ice tray" (action e) are mutex actions in the example illustrated in Figure 3.

4.1.3 Incompatible actions

An action a is incompatible with an action b if b cannot be done before a. For example, "add chemical 4 to mix" (action y) and "take tube (with mix)" (action j) in Figure 3 are incompatible, because if action j has already been performed, action y cannot be performed, since chemicals can only be added to the mix while the tube is on the workbench.

4.1.4 Unordered actions

In some cases, the protocol has groups of actions (also known as compound actions) that can be done in any order. An example of this in Figure 3 is the compound action h, formed by the actions "add chemical 1 to mix" (action h.a) and "add chemical 2 to mix" (action h.b).

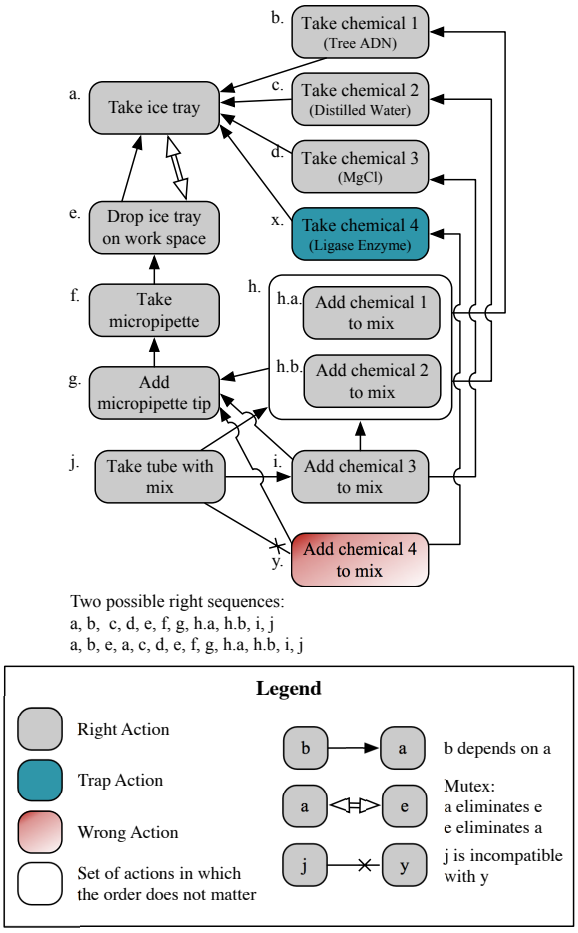


Figure 3: An example of actions and action interdependencies.

4.1.5 Trap actions

A wrong action is an action whose execution constitutes an error in itself. However, the execution of a trap action does not result in an error; it creates the conditions for a student to possibly make a mistake later on. In the practice protocol, action x is a trap action because it is irrelevant for the goals of the practice, but is a prerequisite for a subsequent wrong action (action y).

4.2 Coordination of interactive objects

In OpenSimulator, objects communicate with each other by exchanging messages. Figure 4 shows the sequence of messages that the objects involved in the

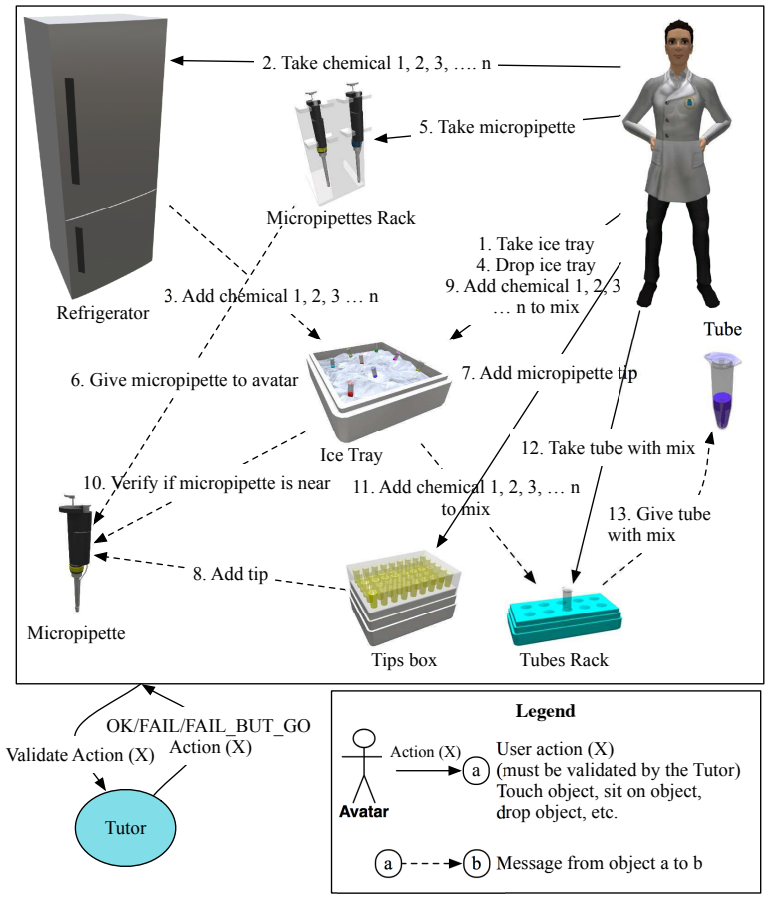


Figure 4: An example of the sequence of messages exchanged between the participant avatar and some of the objects involved in the practice, following the actions and dependencies shown in Figure 3.

above example (specified in Figure 3) exchange with each other while the student is performing the practice. For the sake of clarity, the validation request and the reply of the tutor (OK, FAIL or FAIL BUT GO) have been represented just once at the bottom of the figure, but these two messages are necessary for each user action. If the action is right, the tutor replies OK. Otherwise there are two possible responses by the tutor depending on the configured tutoring strategy: (i) if the reply is FAIL, then the tutor does not give the student permission to perform the action and the student is informed of the reason why; (ii) if the reply is FAIL BUT GO, the tutor does give permission to perform the action, but logs the mistake without immediately warning the student. In this way, students

will personally discover the consequences of their mistakes later on, as they are simulated in the virtual world (in our example, the final result would be a dead plant).

4.3 Tutor configuration

The automatic tutor has been designed to be easily reusable in other virtual laboratories oriented to procedural training implemented on top of OpenSimulator or Second Life (the scripting language code is compatible). In order to reuse the tutor in another virtual laboratory, the most significant task will be to specify the tutor configuration file. This file is divided into several rows where each row is associated with a possible student action. The row number specifies its order in the protocol and its associated tutoring feedback messages. Each row is divided into several fields. The most important fields of a row are as follows:

- OK Message: a message displayed after the current action is executed and validated.
- Next action message: message indicating the next correct action to be executed.
- Dependencies: actions that must have been executed before the current action.
- Incompatibilities: actions that must not have been executed before the current action.
- Locked or unlocked objects: current action may lock (or unlock) some objects that can only be used by one student at a time.
- Max time: action expiration time, i.e., the effects of the action will no longer be felt after this time (in seconds).
- Min time: time (in seconds) that it takes to complete the action.
- Error messages: error messages displayed when current action dependencies, incompatibilities or time constraints are not met.

Table 1 shows an excerpt from the configuration of action i in Figure 3 ("add chemical 3 to mix"). This table outlines the contents of the fields "Dependencies" and "Error messages", providing a link between the dependencies and the associated error messages. In this case, the dependency “(f-g-[h.a-h.b])” can be phrased as "action f, then action g, then action h.a or action h.b". In other words, when a student tries to perform action i, the tutor will display the error message “Chemical 2 has not been added to mix” if he/she forgot to perform action h.b

beforehand. Another possible error is that the student has performed all the required actions but not in the right order. In this case, the tutor will display the error message "Error in the order of the procedure for adding chemical 3 to the mix". As defining the error message configuration for an action, it is possible to specify whether an error message will be shown to the student immediately after validating the action or at the end of the phase. This is indicated by putting 1 or 0 just before the message respectively.

Table 1: Example of dependencies "(f-g-[h.a-h.b])" and error message configuration for action i in Figure 3

Dependencies	**Error messages**
(f	0-You have not taken a micropipette
g	1-You have not attached a micropipette tip
[h.a	0-Chemical 1 has not been added to mix
h.b])	0-Chemical 2 has not been added to mix
Order error	0-Error in the order of the procedure for adding chemical 3 to the mix

An example of the tutor configuration is shown in a spreadsheet in `http://goo.gl/9XrnYg` (messages are in Spanish). Note, however, that, for the sake of clarity and brevity, the example explained above is an adaptation of the real process specified in this spreadsheet.

5 Development

The technical staff of the project was composed of computer science researchers, professors and postgraduate students.

The virtual laboratory development process was addressed iteratively. Accordingly, a phase of the practice was implemented in each iteration. At the beginning of each iteration, the biotechnology expert/professor explained the sequence of actions to the technical staff and discussed how each action of the phase was going to be modeled in terms of avatar actions. Moreover, the technical staff and the expert jointly examined all the possible action sequences required to complete the phase correctly and agreed to add some wrong actions to the virtual laboratory to account for more student errors. During this knowledge acquisition process, the technical staff drew diagrams (see Figure 3), which were validated by the expert. They then specified action dependencies in the

configuration file. Finally, the technical staff and the expert jointly specified the tutoring feedback in the configuration file.

As the implementation of the virtual laboratory progressed, expert participation in action modeling decreased because the technical staff could make more decisions based on their own previous experience.

At the end of each iteration, the prototype was validated by the in-house expert, and many of the mistakes detected in the tutoring feedback could be fixed on the fly by modifying the configuration file at runtime. In this way, the expert was more involved in the pedagogical design of the virtual laboratory. Additionally, by participating in system testing, she gained a better understanding of the limitations and potential of virtual world technology. This was especially helpful in practice modeling.

Concerning the graphics modeling, the in-house expert provided photos and videos of the laboratory objects as well as advised on the most suitable furniture. Some of the objects were modeled by the developers and others were downloaded from online repositories of free objects, mostly, furniture objects.

5.1 Development cost

The virtual laboratory was implemented in 2.5 years. The costs of the main tasks involved in the developments of the virtual laboratory are outlined in Table 2.

Table 2: Development costs of the virtual laboratory

Tasks	Cost (hours)
Expert dedication	92
Development of automatic tutor	624
Firestorm modifications	102
Graphics modeling	150
Other tasks (scripting, testing, etc.)	1180
TOTAL	2149

We should note that the automatic tutor was implemented as a generic component, and may, therefore, be reused in future virtual laboratories for procedural training with the resulting cost savings. In this project, the technical staff also modified the source code of Firestorm viewer (`http://www.firestormviewer.org/`) to simplify its GUI and adapt it to the biotechnology practice.

It took the technical staff around 49 hours to set up the virtual laboratory in the server for a task practice monitoring 25 students. On average, it took

3 hours of a virtual lab administrator's time to monitor a group of three or four students. Taking into account these hours for setting up the server and for monitoring students, and the total cost of developing the virtual laboratory, we estimate that the task practice has a cost of around 800 euros per student. This cost has been calculated considering that only 62 students have performed the task practice so far. Therefore, the cost per student will be reduced in the future.

6 Expert Evaluation

A group of 15 experts in biotechnology helped us to assess the quality of the virtual laboratory. The number of participants was established as a trade-off between precision and effort. In order to achieve a 90% confidence level for a given mean with an error of less than 1%, at least 15 measurements have to be taken [Efron and Tibshirani 1986]. This assumes that intervals are based on a normal population distribution for the sample mean, as is the case here.

These experts are members of the Center of Biotechnology and Plant Genomics at the UPM (our institution). Their demographic characteristics are: age range of 30-45 years, 70% female. Although small, this group is a significant representation because the participants are experts in the subjects covered by this virtual laboratory and they all have teaching experience.

Participants played the student role and were enrolled in the virtual practice. They performed a use case covering most of the functionality described in the previous sections. Although there was no maximum time to complete the practice, it took most users around 60 minutes.

At the end of the practice, a detailed questionnaire was administered to each participant (available at `http://sites.google.com/site/biotechquestionnaire` under the "Resources" link in the top left-hand corner of the site). This questionnaire has 50 questions evaluating three different features: (a) usability of the virtual laboratory, (c) user levels of agreement/disagreement with the educational value of the tool and (c) user satisfaction concerning the virtual laboratory user interface. Parts a and c are based on standard questionnaires, described below, and part b is an ad hoc questionnaire.

Tables 3, 4 and 5 show the average values and standard deviations of this questionnaire. The following sections provide a detailed description of each part and the results found.

6.1 Usability evaluation

The usability of the system was measured by means of a well-established standard test named "Practical Heuristics for Usability Evaluation" [Perlman 1997]. This test includes 13 questions ranging from 1 (bad) to 5 (good), providing a useful measure of user-perceived usability. This set of questions is divided into

Table 3: Satisfaction part of the questionnaire (experts)

Question	Mean	Standard Deviation
S-O1	6.3	0.5
S-O2	4.7	1.2
S-O3	5.1	1
S-O4	5.6	1.1
S-O5	5.9	0.8
S-O6	4.5	1.3
S-P1	6	1.3
S-P2	5	1.5
S-P3	5.3	1.6
S-I1	6.2	0.8
S-I2	4.3	1.7
S-I3	4.9	1.6
S-I4	5.2	1.9
S-I5	4.3	2.1
S-L1	5.7	1.4
S-L2	5.1	1.5
S-L3	5.7	1
S-L4	5.1	1.2
S-L5	5.5	1.2
S-L6	5.8	1.3
S-C1	4.9	1.7
S-C2	5.1	1.6
S-C3	4.8	1.5
S-C4	4.5	1.7
S-C5	5.4	1.5
Mean	5.3	1.3

Label: Min values Max values

three blocks: the first four questions are related to "learning" (U-L1 to U-L4), the following five are related to "adapting to the user" (U-A1 to U-A5) and the final four are related to "feedback and errors" (U-F1 to U-F4). Two additional questions are open-ended questions regarding positive or negative aspects of the virtual laboratory and its usage.

The results of this test are shown in Table 4. The high values in the "learning" block and "adapting to the user" block of the user evaluation indicate that experts are satisfied by the help and documentation available in the virtual laboratory

Table 4: Usability part of the questionnaire (experts)

Question	Mean	Standard Deviation
U-L1	4.6	0.5
U-L2	4.2	0.9
U-L3	4.2	1
U-L4	4.3	0.7
U-A1	4.3	0.6
U-A2	3.4	1
U-A3	4.8	0.4
U-A4	4.5	0.6
U-A5	4.5	0.5
U-F1	4.4	0.9
U-F2	3.7	1
U-F3	3.9	0.9
U-F4	3.1	1
Mean	4.1	0.8

Label: Min values Max values

Table 5: Educational part of the questionnaire (experts)

Question	Mean	Standard Deviation
E-1	4.5	0.7
E-2	4.8	0.4
E-3	4.1	0.9
E-4	3.8	1
E-5	4.7	0.6
E-6	4.8	0.4
Mean	4.5	0.7

Label: Min values Max values

and the quality of the graphical objects.

On the other hand, as shown by the lower values in the "feedback and errors" block especially, experts are less satisfied with the automatic tutor's instructions about what to do at any time or how to return to a correct state after an error.

In the open-ended questions, experts reported some problems with avatar handling and that they found the initial steps of the practice hard to understand.

6.2 Evaluation of the educational value perceived by users

As Table 5 (with integer values in the 1-5 range) shows, participants are enthusiastic about the educational potential of the virtual laboratory. In particular, experts drew attention to the fact that the virtual practice was able to reinforce student knowledge (E-2) and stated an interest in having a similar virtual laboratory for other subjects (E-6).

6.3 User satisfaction evaluation

In order to evaluate user satisfaction, we used a slightly modified version of the standard "User Interface Satisfaction" test [Chin et al. 1988]. The standard version includes 27 questions, but was reduced to 25 due to overlaps with the usability test described above.

Valid responses to these questions are positive integers ranging from 1 (not at all satisfied) to 7 (completely satisfied). The results are shown in Table 3.

These questions were divided into six blocks: overall impression of the laboratory (S-O1 to S-O6), presentation on screen (S-P1 to S-P3), terminology and system information (S-I1 to S-I5), learning (S-L1 to S-L6) and system capabilities (S-C1 to S-C5).

Overall the experts' impression of the laboratory was positive (S-O1). Besides, their rating of the use of the biotechnology terms throughout system was high. On the other hand, some experts were not completely satisfied with the position of messages on screen and the system's error correction capabilities.

7 Student Evaluation

Following on from the expert evaluation, we conducted a student evaluation where they were asked to perform a virtual laboratory practice individually and fill in a questionnaire very like the survey completed by the experts.

The experiment was run in the second semester of 2014, 2015 and 2016 with 31, 15 and 16 students of the BSc in Forestry Engineering and BSc in Natural Environment Engineering, respectively, enrolled in the "Biochemistry and Biotechnology" core subject, which is compulsory at UPM. Their demographic characteristics are: age range of 18-19 years and around 30% female. The practice was optional, but students performing the practice received a bonus in the final course grade.

At the beginning of the study, none of the students had any previous experience of working in a real biotechnology laboratory. Immediately before they did the practice, they received a two-hour lecture with slides on biotechnology concepts related to the practice and its protocol. However, this lecture did not

cover all the information required to perform the practice. This did not prevent students from successfully completing the practice, because all the missing information was available within the virtual laboratory.

Afterwards, one of the researchers gave students a one-hour talk on the procedure of the study and the basics of avatar control, and supplied a help card on avatar control. Then, students had to install the client application on their own computers and connect to the virtual laboratory to complete a brief tutorial on their own. After performing the tutorial, they were ready to start the biotechnology practice. The virtual practice covered around 120 basic actions grouped into three phases. It took students around 90 minutes on average to complete the entire practice. At first glance, it is perhaps surprising that there is not a bigger difference between the times taken by experts (60 minutes on average) and students. A possible explanation, however, is that experts did not receive any previous training on avatar handling, although they were trained by researchers while they completed the practice. Additionally, each student was assisted throughout the practice by the automatic tutor, as explained in Section 4. This helped to reduce the gap between experts and students. Students were asked to use their own computers and Internet connections to perform the practice without the physical presence of technical staff or a professor; however, an avatar controlled by a member of the technical staff was always on hand in the virtual laboratory to help students via chat with any issue related to avatar control or object location. Apart from the instructions provided by the automatic tutor, students could also consult some slides within the virtual world with a more detailed description of the practice protocol than the one provided in the slides presented in the previous lecture.

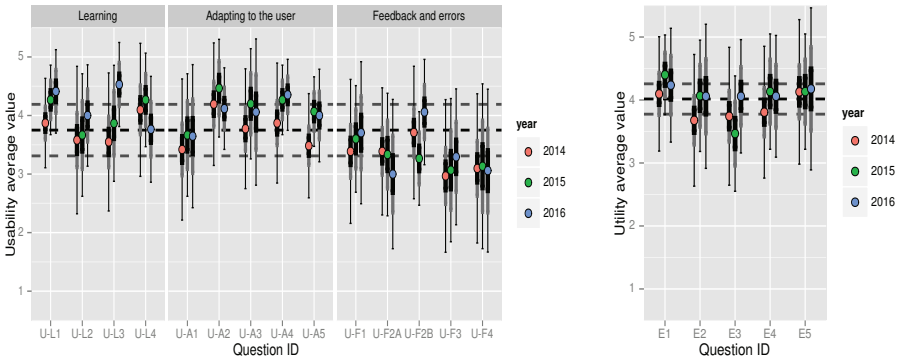
Immediately after completing the practice, students had to fill in the questionnaire on the virtual laboratory. This questionnaire was the same as the one completed by the professors, except for some minor points introduced to improve the clarity of some questions or to collect more specific information on some issues.

Figure 5 shows that most values tend to be better in 2015 and 2016 than in 2014. This is because some bugs were detected in the virtual laboratory and fixed in the first round. Therefore, these bugs did not detract from the user experience of students in 2015 and 2016.

Tables 6, 7 and 8 report the aggregated data for these three years.

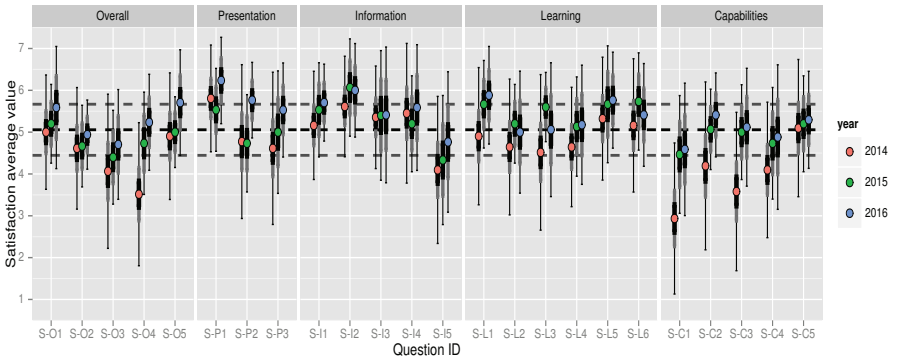
7.1 Usability evaluation

From results in Table 7 we can conclude that students agree on the learning benefits of the system, but they miss more feedback and error information. Besides, many students agree that graphical objects are a faithful representation of their use and look in the real world. Additionally, many students also highlight the



(a) Usability

(b) Educational



(c) Satisfaction

Figure 5: Mean value of student questionnaire responses (circles) for the three parts of the questionnaire. The thin bars in each circle show standard deviation. The thick bars show the 90% confidence interval of the mean. Black dotted lines show the average value of the responses, and grey dotted lines show standard deviation bounds.

ease of use for a user with basic computer skills. This suggests that interaction with the virtual laboratory was no obstacle to most students.

The poorest results were for the Feedback and errors category. Above all, these results suggest that there is room for improvement of the automatic guidance provided when students make a mistake.

If we compare the results of the experts and students, most of the questions are, as expected, rated higher in the expert evaluation. However, we would like to point out some interesting differences, especially in the learning block. Firstly,

Table 6: Satisfaction part of the questionnaire (students)

Question	Mean	Standard Deviation
S-O1	5.2	1.3
S-O2	4.7	1.1
S-O3	4.3	1.4
S-O4	4.3	1.3
S-O5	5.1	1.2
S-P1	5.9	1.1
S-P2	5	1.3
S-P3	5	1.4
S-I1	5.4	1.1
S-I2	5.8	1.2
S-I3	5.4	1.5
S-I4	5.4	1.4
S-I5	4.3	1.7
S-L1	5.3	1.3
S-L2	4.9	1.3
S-L3	4.9	1.4
S-L4	4.9	1.3
S-L5	5.5	1.3
S-L6	5.4	1.3
S-C1	3.7	1.6
S-C2	4.7	1.3
S-C3	4.3	1.4
S-C4	4.5	1.6
S-C5	5.2	1.3
Mean	5	1.3

Label: Min values Max values

experts have a rather better opinion of the automatic tutor than students regarding what to do in the next step. We think that a possible explanation is that, whereas experts consider the assistance provided by the tutor to be sufficient taking into account learning objectives, students would prefer to receive more help. Also, due to the implemented educational strategy, the tutor allows students to make mistakes, and its hints are, intentionally, not always detailed enough. This is not always well received by the students, who feel that the tutor does not provide them with solutions to their problems. In the Feedback and errors block, it is noteworthy that experts are rather more satisfied with

Table 7: Usability part of the questionnaire (students)

Question	Mean	Standard Deviation
U-L1	4.1	0.7
U-L2	3.7	1
U-L3	3.9	0.9
U-L4	4	0.9
U-A1	3.5	1.1
U-A2	4.2	0.9
U-A3	4	1
U-A4	4	0.7
U-A5	3.8	0.8
U-F1	3.5	1.1
U-F2A	3.3	1.1
U-F2B	3.7	0.9
U-F3	3.1	1.2
U-F4	3.1	1.4
Mean	3.7	1

Table 8: Educational part of the questionnaire (students)

Question	Mean	Standard Deviation
E-1	4.2	0.8
E-2	3.9	1
E-3	3.8	1
E-4	4	1
E-5	4.1	1.1
Mean	4	1

Label: Min values Max values

the feedback on what is happening. The difference on this point is justified by the fact that experts are more experienced. This provides them with a deeper comprehension of the situation in not only the real but also in the virtual world.

On the other hand, it is worth mentioning that both experts and students score error correction instructions lowest (see U-F3 and U-F4 results). We think that these low ratings are due to the fact that the virtual laboratory is not designed to support undo wrong actions or to give more specific instructions when students fail to take the right action after several attempts.

7.2 Evaluation of the educational value perceived by users

From the results in Table 8, we can conclude that students have a good opinion of the educational value of the virtual laboratory. For example, they rate the correspondence between the virtual laboratory and what was explained in the classroom positively (E-1).

Comparing the results between experts and students, we find that, generally speaking, experts had a better opinion of the educational value of the virtual laboratory. The difference for question E-2 is particularly noteworthy. This may be due to the fact that some students considered that the new things that they learned by doing the practice were not a reinforcement of something previously studied in class. On the other hand, both experts and students rated the use of a similar virtual laboratory for other subjects highly.

7.3 User satisfaction evaluation

As Table 6 shows, the best results are for the Information section. This suggests that students like the message presentation. Nevertheless, message sequencing and organization do not get such good results as message presentation. In contrast, question S-L5 (help messages in the range of unhelpful to helpful) score quite highly. Clearly, students score better the hints for right actions than the interaction associated with their errors.

Questions S-O4 and S-C1 are the lowest scorers. Although they belong to different sections, they are related to the system time performance. We think that these values are the result of lags in the reception of messages from the tutor occasionally experienced by students during the practice. We will address this issue in more detail in the Discussion section.

If we compare the results between experts and students with respect to satisfaction, we find that students are more critical than experts of the automatic tutor lags (S-O4 and S-C1). Instead, students scored questions related to terminology and system information higher (S-I1 to S-I5).

8 Discussion

8.1 Discussion of the pedagogical approach

The results of the experiment show that the evaluated virtual laboratory was positively received by experts and students. However, the experiment revealed some issues with the virtual laboratory that might be improved. One of the main aspects that would require further enhancement in the opinion of the users is student error management. In this respect, users would like to have the option of repeating some parts of the practice. This should be weighed against whether it is

possible to do so in the real world and pedagogically suitable. Also, after making a mistake, some users need more detailed assistance from the automatic tutor on how to do things correctly. This may involve an explanation of the reasons for the mistake. Thus there is a need to provide more personalized tutoring for any students requiring help.

Another possible enhancement of the virtual laboratory, according to the opinion of the users, would be to give users more freedom to explore the consequences of taking actions beyond the pre-established practice protocol. Such extensions would logically increase the development cost of the virtual laboratory and should be carefully selected by the professor so as not to detract from the learning objectives of the practice.

8.2 Discussion of the technical approach

Concerning the virtual world platform, we consider that OpenSimulator (OS) and Second Life are a good option for the development of many virtual laboratories with an acceptable graphical quality. However, the lags observed in the automatic tutor responses point to a system bottleneck, which may be solved by reimplementing the automatic tutor. The automatic tutor uses listeners to communicate with the other objects of the virtual laboratory, which, as is explained in [Weber et al. 2007], significantly degrade performance on Second Life and very presumably OpenSimulator. Additionally, these script languages do not support pointers. Therefore, efficient dynamic data structures cannot be implemented. As result, we may achieve a significant performance improvement by reimplementing the automatic tutor in C# (the server implementation language), that is, moving the script code related to the tutor to server methods. In our first performance tests with a prototype tutor implemented in C#, we have found that the new tutor is at least 15 times faster than the version in the OS script language. Another advantage of implementing the tutor in C# would be that the tutor may be reused in virtual laboratories implemented on top of other graphic engines, such as Unity 3D.

The results of the experiments indicate that interaction mechanisms were not an obstacle for most students.

As mentioned above, the total costs per student for the virtual laboratory (see section 5.1, 800 euros) are quite lower than the costs (see section 3, between 5,000 and 25,000 euros) for a real laboratory. Additionally, if we also take into account the positive results of the evaluation, all this evidence supports the claim that the virtual laboratory is cost-effective.

As the virtual laboratory has been used by 62 students in the evaluation, the resilience and the reliability of the virtual laboratory have been demonstrated.

9 Conclusion

This paper presents an empirical evaluation of a VWPT with domain experts and students. The analysis of this evaluation has shown good results for usability, user satisfaction and perceived educational utility. To the best of our knowledge, this is one of the few VWPTs equipped with an automatic tutor that has been evaluated by domain experts and students of a university degree course.

From the perspective of the development process, we argue that this project represents an real example of how a 3D virtual laboratory can be built at an affordable cost for educational institutions or companies. In addition, the evaluation shows the resilience and the reliability of the virtual laboratory and, in particular, the automatic tutor.

One of the main contributions of this project is the automatic tutor. The automatic tutor was designed as a generic component to be reusable in future virtual laboratories for procedural training developed on top of OpenSimulator or Second Life.

10 Future Work

We expect to alleviate the performance problems detected in the evaluation by reimplementing the automatic tutor in C#. Additionally, we will study how to modify the design of the automatic tutor to be multithreaded. In this way, we will improve its scalability, because it will be able to leverage multiple processors when they are available and the number of students requires it.

Currently, we are researching how to leverage student logs to enhance the tutoring feedback [Riofrío and J. Ramírez 2015]. Another approach that we are researching is more focused on customization of feedback based on student modeling [Clemente et al. 2011]. Through such customization, we would be able to implement scaffolding techniques. The provided support should be tailored to each student's zone of proximal development (ZPD) [Vygotsky 1978].

Acknowledgment

We would like to acknowledge funding from the UPM Educational Innovation Program (Projects IE 100210147, PT 07017-1112 and IE 12-1307001) and thank GIE-LVBA members. Riofrío received funding from the Secretariat of Higher Education, Science, Technology and Innovation of Ecuador (SENESCYT), and Rico acknowledges support from projects TIN2011-24139 and JCI-2012-12719. Thanks to Rachel Elliot for her comments on the paper.

References

[Buche et al. 2003] Buche, C., Querrec, R., De Loor, P.: "MASCARET: Pedagogical Multi-Agents Systems for Virtual Environment for Training"; Cyberworlds, (2003).

[Brown et al. 2012] Brown, R., Rasmussen, R., Baldwin, I., Wyeth, P.: "Design and Implementation of a Virtual World Training Simulation of ICU First Hour Handover Processes"; Australian Critical Care : Official Journal of the Confederation of Australian Critical Care Nurses, 25 (Aug 2012), 178-87.

[Chin et al. 1988] Chin, J. P., Diehl, V. A., Norman, K. L.: "Development of an Instrument Measuring User Satisfaction of the Human-computer Interface"; Proc. SIGCHI Conference on Human Factors in Computing Systems, ACM, New York (1988), 213-218.

[Clemente et al. 2011] Clemente, J., Ramírez, J., de Antonio, A.: "A Proposal for Student Modeling Based on Ontologies and Diagnosis Rules"; Expert Systems with Applications, 38, (Jul 2011), 8066-8078.

[Dalgarno 2002] Dalgarno, B.: "The Potential of 3D Virtual Learning Environments: A Constructivist Analysis"; Electronic Journal of Instructional Science and Technology (2002), 1-19.

[Dalgarno and Lee 2012] Dalgarno, B., Lee, M. J. W.: "Exploring the relationship between afforded learning tasks and learning benefits in 3D virtual learning environments"; Proc. ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference, (2012).

[Duncan et al. 2012] Duncan, I., Miller, A., Jiang, S.: "A taxonomy of virtual worlds usage in education"; British Journal of Educational Technology, 43, 6 (2012), 949-964.

[Efron and Tibshirani 1986] Efron, B., Tibshirani, R.: "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy"; Statistical science, (1986), 54-75.

[Faria et al. 2009] Faria, L., Silva, A., Ramos, C., Vale, Z., Marques, A.: "Cyber-Ambient Intelligent Training of Operators in Power Systems Control Centres"; Proc. 15th International Conference on Intelligent System Applications to Power Systems, (2009).

[Fletcher. 1998] Fletcher, B.: "TRANSoM ROV Pilot Training: Test Results"; Underwater Intervention '98 Conference Proceedings, (1998).

[Gu et al. 2015] Gu, Y., Sosnovsky, S., Ullrich, C.: "SafeChild: An Intelligent Virtual Reality Environment for Training Pedestrian Safety Skills"; Design for Teaching and Learning in a Networked World SE - 11, Lecture Notes in Computer Science, Springer International Publishing (2015), 141-154.

[Hill et al. 2006] Hill, R., Gratch, J., Johnson, W. L., Kyriakakis, C., LaBore, C., Lindheim, R., Marsella, S., Miraglia, D., Moore, B., Morie, J., Rickel, J., Thiébaux, M., Tuch, L., Whitney, R., Douglas, J., Swartout, W.: "Toward the Holodeck: Integrating Graphics, Sound, Character and Story"; Proc. Fifth International Conference on Autonomous Agents, ACM, (2001), 409-416.

[Holden and Sinatra 2014] Holden, H. K., Sinatra, A. M.: "A Guide to Scaffolding and Guided Instructional Strategies for ITSs"; Design Recommendations for Intelligent Tutoring Systems, Orlando: U.S. Army Research Laboratory, (2014), 265-281.

[İnce et al. 2014] İnce, E., Kırbaşlar, F. G., Yolcu, E., Aslan, A. E., Kayacan, Z. C., Alkan Olsson, J., Akbasli, A. C., Aytekin, M., Bauer, T., Charalambis, D.: "3-Dimensional and Interactive Istanbul University Virtual Laboratory Based on Active Learning Methods."; Turkish Online Journal of Educational Technology-TOJET, 13, 1 (2014), 1-20.

[İnce et al. 2015] İnce, E., Kırbaşlar, F. G., Güneş, Z. Ö., Yaman, Y., Yolcu, Ö., Yolcu, E.: "An Innovative Approach in Virtual Laboratory Education: The Case of "IU-VIRLAB" and Relationships between Communication Skills with the Usage of IU-VIRLAB"; Procedia - Social and Behavioral Sciences, 195, (Jul 2015), 1768-1777.

[Johnson and Rickel 1997] Johnson, W. L., Rickel, J.: "Steve: An Animated Pedagogical Agent for Procedural Training in Virtual Environments"; SIGART Bulletin, 8, (1997), 1-4.

[Kapur 2008] Kapur, M.: "Productive Failure; Cognition and Instruction, 26, 3 (2008), 379-424.

[Loftin and Kenney 1995] Loftin, R., Kenney, P.: "Training the Hubble space telescope flight team.';' IEEE Computer Graphics and Applications, (1995).

[Los Arcos et al. 2000] Los Arcos, J. L., Müller, W., Fuente, O., Orúe, L., Arroyo, E., Leaznibarrutia, I., Santander, J.: "LAHYSTOTRAIN: Integration of Virtual Environments and ITS for Surgery Training"; Intelligent Tutoring Systems SE, Lecture Notes in Computer Science, 1839, (2000), 43-52.

[Merriam 1998] Merriam, S. B.: "Qualitative Research and Case Study Applications in Education. Revised and Expanded from "Case Study Research in Education""; ERIC, (1998).

[Messienger et al. 2009] Messinger, P. R., Stroulia, E., Lyons, K., Bone, M., Niu, R. H., Smirnov, K., Perelgut, S.: "Virtual Worlds - Past, Present, and Future: New Directions in Social Computing"; Decision Support Systems, 47, 3 (2009), 204-228.

[Olney 2014] Olney, A. M.: "Scaffolding Made Visible"; Design Recommendations for Intelligent Tutoring Systems, Orlando: U.S. Army Research Laboratory, (2014), 327-340.

[Perlman 1997] Perlman, G.: "Practical usability evaluation"; CHI'97 Extended Abstracts on Human Factors in Computing Systems, ACM, (1997), 168-169.

[Pioch et al. 1997] Pioch, N. J., Roberts, B., Zeltzer, D.: "A Virtual Environment for Learning to Pilot Remotely Operated Vehicles"; Proc. International Conference on Virtual Systems and MultiMedia VSMM '97, (1997).

[Rico et al. 2012] Rico, M., Ramírez, J., Riofrío, D., Berrocal-Lobo, M., De Antonio, A.: "An Architecture for Virtual Labs in Engineering Education"; Proc. Global Engineering Education Conference (EDUCON), IEEE, Marrakesh (2012), 1-5.

[Riofrío and J. Ramírez 2015] Riofrío, D., Ramírez, J.: "A Model for Student Action Prediction in 3D Virtual Environments for Procedural Training"; Proc. Educational Data Mining 2015: 8th International Conference on Educational Data Mining, Madrid (2015).

[Rudman et al. 2010] Rudman, P. D., Lavelle, S. P., Salmon, G., Cashmore, A.: "SWIFT-ly Enhancing Laboratory Learning: Genetics in the Virtual World"; Proc. Association of Learning Technology, ALT-C, (2010).

[Vygotsky 1978] Vygotsky, L. S.: "Mind in Society: The Development of Higher Psychological Processes", Harvard University Press, (1978).

[Weber et al. 2007] Weber, A., Rufer-Bach, K., Platel, R.: "Creating your world: The official guide to advanced content creation for Second Life"; John Wiley & Sons, (2007), 370.

[Zhong and Liu 2012] Zhong, Y., Liu, C.: "User Evaluation of a Domain-Oriented End-User Design Environment for Building 3D Virtual Chemistry Experiments"; User Evaluation for Software Engineering Researchers (USER), IEEE, (2012).