Students' Evaluation of a Virtual World for Procedural Training in a Tertiary-Education Course

**Jaime Ramírez¹, Mariano Rico²,
Diego Riofrío-Luzcando¹, Marta Berrocal-Lobo³,
and Angélica de Antonio¹**

Abstract

This article presents an investigation on the educational value of virtual worlds intended for the acquisition of procedural knowledge. This investigation takes as a case of study a virtual laboratory on biotechnology. A remarkable feature in this virtual laboratory is an automatic tutor that supervises student's actions and provides tutoring feedback when it is appropriate. The study presented in this article covers two different aspects of the system. First, it analyzes the impact of this virtual world in learning some concepts related to a biotechnology practice; and second, it surveys the students' opinion on the virtual world by means of three open questions (positive, negative, and general impressions). Results demonstrated that the virtual world had a positive influence in the students' knowledge, and it was well received by them.

¹School of Computer Engineering, UPM, Spain

²Department of Artificial Intelligence, UPM, Spain

³MONTES (School of Forest Engineering and Natural Environment), UPM, Spain

Corresponding Author:

Jaime Ramírez, Escuela Sup. de Ing. Informáticos, Calle de los Ciruelos Montegancedo Campus, Universidad Politécnica de Madrid, Building 5, Floor 1, Office 5112, 28660 Boadilla del Monte, Madrid, Spain.

Email: jramirez@fi.upm.es

There are many examples in the literature that show how interactive simulations are successfully employed for educational purposes. Among these interactive simulations, three-dimensional (3D) virtual environments represent a group of applications that each day is attracting more and more the interest of the educational community. This is because these systems are engaging and allow students to learn new concepts and procedures by recreating situations that in the real world would be too expensive, dangerous, or simply out of reach for the students. We can see 3D virtual environments as a tool for implementing the “learning by doing” approach derived from the constructivist pedagogy (Huang, Rauch, & Liaw, 2010).

When several users can visit a shared 3D virtual environment through Internet and interact by means of desktop devices, this 3D virtual environment is usually referred to as a virtual world. In addition to the benefits of 3D simulations, virtual worlds also provide the possibility of reproducing real scenarios where several people (students or teachers) are able to chat and share rooms, instruments, and so forth. Moreover, as students connect to virtual worlds through Internet, they can perform virtual practices at home, without the physical presence of a teacher close to them. This opens the door to virtual practices from distant places where sufficiently equipped real labs are not available.

In recent years, the development of 3D virtual worlds has become more inexpensive thanks to platforms such as Second Life (<http://secondlife.com/>), OpenSimulator (OS; <http://opensimulator.org/>), Open Wonderland (<http://openwonderland.org/>), or Bitmanagement Software Collaborate System (<http://www.bitmanagement.com/>), which provide developers with a basic layer of functionality. As a result of this, more academic institutions have been able to develop and use their own virtual worlds.

This article presents a study on the educational value of a virtual world for procedural learning. In the literature on virtual worlds, we can find many works on how this kind of systems can support different learning activities. However, there are just a few works that show how this kind of systems can support task practices effectively. We believe it is so because virtual worlds for procedural training are more complex than other educational virtual worlds, especially when they incorporate automatic tutoring.

To explore the educational value of virtual worlds for procedural training, we have implemented and validated a good representative of this family of systems. We believe that most of our conclusions are also applicable to other virtual worlds for procedural training.

The chosen case study was a virtual laboratory of biotechnology (VLB). The empirical evaluation has been carried out with first-year students of the agroforestry engineering school at Universidad Politécnica de Madrid (UPM—Spain). The experiments were conducted along 2 academic years with students enrolled in the subject “Biochemistry and Biotechnology.”

This study covers two different aspects of the system. First, it analyzes the impact of the virtual world in the learning of some concepts related to a biotechnology practice; and second, it surveys the students' opinion on the virtual world by means of three open questions (positive, negative, and general impressions). As we will show in this article, the virtual world had a positive influence in the students' knowledge, and it was well received by them.

Our system implements an automatic tutor that supervises the student's behavior and provides tutoring feedback in the form of error messages and hints, whenever it is pedagogically appropriate. As we will show in the next section, even if there are some studies in the literature that justify with empirical results the educational value of 3D virtual environments, only a few works report empirical results of a system equipped with an automatic tutor. To the best of our knowledge, there is no other work that has conducted a similar study, evaluating the usefulness of a virtual world for learning procedures (or protocols), with the help of an automatic tutor, at a higher education course.

The structure of the remainder of the article is as follows. The next section presents a survey of 3D virtual environments oriented to education with a particular focus on those that have reported empirical evaluations. Next, we briefly explain the main features of our virtual lab. Thereupon, the article outlines the details of the study, explaining its goals, the design of the experiments, the procedure that was followed, and the instruments employed; next, a qualitative analysis of the collected data is exposed, followed by the interpretation of the results. Finally, some general conclusions and future lines of research that may be derived from this work are presented.

Related Work

In the software community, there exists a wide variety of educational virtual environments. Some of them show information to the students through pictures, videos, interactive objects, or they help teachers to make virtual lectures. These educational virtual environments can be employed either as an alternative to traditional learning methods or as a supplementary method.

However, virtual environments will not support an effective learning unless they have some specific characteristics, such as the principles proposed by Clark and Mayer (2011) for educational games and simulators to facilitate learning: match the game or simulation type to learning goals, make learning essential to progress in the game or simulation, use of well-proven instructional strategies (explanatory feedback, pretraining, etc.), offer learning support (guidance, visualization support, instructional explanations, etc.) and a solid domain-specific structure, manage the complexity of the game or simulation depending on the goals and learners' expertise (easy to use, align pace to instructional goals, etc.), and make the learning relevancy salient to the students. Clark and Mayer (2011) for educational games and simulators to facilitate learning: match the game

or simulation type to learning goals, make learning essential to progress in the game or simulation, use of well-proven instructional strategies (explanatory feedback, pretraining, etc.), offer learning support (guidance, visualization support, instructional explanations, etc.) and a solid domain-specific structure, manage the complexity of the game or simulation depending on the goals and learners' expertise (easy to use, align pace to instructional goals, etc.), and make the learning relevancy salient to the students.

Holzinger, Kickmeier-Rust, Wassertheurer, and Hessinger (2009) showed with an experiment that interacting with sophisticated simulations strain the limitations of cognitive processes; therefore, successful application of simulations requires careful additional guidance from tutors and a certain amount of previous knowledge on the part of the learners. In another experiment, Song et al. (2014) demonstrated that a high cognitive interactivity in learning environments can harm learning effectiveness; therefore, instructional designers must consider cognitive load in interaction design. These results motivate the integration of automatic tutors with virtual environments because they can assist students in learning tasks by reducing their extraneous cognitive load.

According to Dalgarno and Lee (2012), educational virtual worlds can be employed to support the following learning activities (not mutually exclusive): place exploration, concept exploration, task practice, role-play, gaming, communication, slide show, building or scripting, instruction, and machinima. Among these learning activities, we will focus on how virtual worlds can support task practice successfully. In task practice, students have to exercise procedural tasks with the goal of learning how to perform these same tasks in the real world. In many cases, these tasks can be dangerous, overly expensive or too time consuming to be practiced in the real world. Moreover, while students practice a task, they can also learn new concepts or reinforce already learned ones. Virtual laboratories are typical examples of virtual worlds where students can perform task practice. In addition, we will constrain our survey to training in closed-ended tasks, in which there is only one procedure to achieve a valid final state, or more than one similar procedure with minor differences among them.

Educational 3D Virtual Environments for Procedural Training

In this section, besides covering virtual worlds for procedural training (VWPT), we will expand our scope to other kinds of 3D virtual environments for procedural training (VEPT) that do not have multiuser support or that rely on haptic devices, but which pose some features of interest for our study. Typically, the usage of haptic devices is associated with the learning of some manual skills. However, we will not address the implications of using haptic devices, since we focus on procedural training.

The use of 3D virtual environments 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 domain knowledge, students discover a deficiency in their own knowledge or an inconsistency between their representation of knowledge and their experience. Moreover, also according to this view, learning occurs in a social context in which the interaction between students and their peers is a necessary part of the learning process.

It is worth highlighting that the work in VEPTs is scarce in comparison with the work in web applications for procedural training where we can find many systems, even equipped with automatic tutors. A representative sample of this latter kind of applications includes INES, Adele, Assistment, and “Cognitive Tutor,” described later.

INES (Hospers, Kroezen, Nijholt, op den Akker, & Heylen, 2003) is a multi-agent system for nursing education, which provides explanations or demos to students when they make mistakes.

Adele (Johnson, Shaw, Marshall, & LaBore, 2003; Shaw, Johnson, & Ganeshan, 1999) teaches how to do medical diagnosis and treatments relying on an automatic tutor that can provide hints and explanations on demand. Adele has been evaluated with satisfaction questionnaires demonstrating a wide acceptance by students.

In the field of teaching math, we can mention Assistment (Razzaq et al., 2005), which implements a scaffolding strategy to guide students by means of multiple choice questions. An analysis with data collected from several students concluded that students that received scaffolding improved more than those who only received hints. Another tool for learning math is Cognitive Tutor (Ritter, Anderson, Koedinger, & Corbett, 2007; Ritter, Carlson, Sandbothe, & Fancsali, 2015), a very popular system used by hundreds of thousands of students, and whose automatic tutor, according to an empirical study, has proved to be able to increase the students’ grades.

One possible explanation of why the work in VEPTs is scarce compared with the work in web-based procedural training is that, in a VEPT, interaction possibilities are quite richer than in web applications, which makes the development of a VEPT more expensive and the design of an automatic tutor more complex. Another aspect that encourages the development of web applications is the ease to access them because the student only needs a web browser. Nevertheless, we think that the current maturity of VWPT platforms will help to reduce this gap between VWPTs and web applications in the future.

Concerning VWPTs, one of the most well-known and complete systems is Soar Training Expert for Virtual Environments (Johnson & Rickel, 1997), a 3D animated agent designed to assist one or more students to learn a task that follows a given procedure. Another well-known VWPT is MASCARET (Buche, Querrec, & De Loor, 2003), which is an agent-based system that integrates an intelligent tutoring system (ITS; Sleeman & Brown, 1982) and was used to create virtual training environments like Sécuvéi. Although all of these works

represent recognized contributions to the state of the art, they do not report empirical evaluations. Next, we will focus on those works that have actually reported some type of empirical evaluation.

We will begin with studies related to VEPTs that are not equipped with an automatic tutor. A few of these studies have evaluated the educational value of VEPTs with satisfactory results, like LapSim and DentSim, where the group of students trained with the system obtained better results than the control group. LapSim (Larsen et al., 2009) is a system that reproduces many scenarios for laparoscopic surgery training using haptic or nonhaptic controllers to move simulated instruments. DentSim (LeBlanc, Urbankova, & Hadavi, 2004) is another haptic simulator that consists of a patient mannequin with a set of teeth, rotary dental instruments, and a monitor which reproduces a 3D simulated image of the patient's mouth, where users can observe any cut made in the teeth of the mannequin.

Other studies have evaluated whether VWPTs were well accepted or not by the users, such as iVirtualWorld (Zhong, 2013; Zhong & Liu, 2012), which is more than a virtual lab, since it is a tool that allows professors to create their own chemical experiments as a virtual lab. Most of the participants appreciated the ease-of-learn, the ease-of-use, and the usefulness of iVirtualWorld. Another recent work that evaluated the satisfaction of the users was SafeChild (Gu, Sosnovsky, & Ullrich, 2015), which is a virtual city environment with an ITS that provides children a guided and assisted training in pedestrian safety. Although the conclusions of the experiment were positive, they only evaluated the virtual environment without the assistance of the ITS, that is, they focused on the simulation and how children performed the pedestrian exercises.

University virtual laboratory (İnce et al., 2014, 2015) is a VWPT that implements a Physics virtual lab concerning the topics of magnetic fields and magnetism. This work presents a different kind of empirical study, where they focused on analyzing the capacity of the system to improve the communication skills of students.

Two VEPTs that integrate an automatic tutor and have been empirically evaluated are Lahystotrain and CanadarmTutor. Their evaluation relied on usability and satisfaction questionnaires, and the conclusions were especially satisfactory regarding tutoring feedback. Lahystotrain (Los Arcos et al., 2000; Müller-Wittig et al., 2000) was developed to train surgeons in laparoscopy and hysteroscopy operations. This application contains an ITS that sends proactive and reactive help on how to properly perform a surgery of this type to students in a virtual environment. CanadarmTutor Tutoring System (Fournier-Viger, Nkambou, Nguifo, Mayers, & Faghihi, 2013) is a simulator of Canadarm2, a seven-degrees-of-freedom robotic arm used in the International Space Station. This simulator trains astronauts on how to operate this robotic arm. CanadarmTutor provides help to users on how to follow a correct solution

path with hints or generating demonstrations of arm operations. This ITS includes a cognitive model to assess skills and spatial reasoning and an expert system that automatically generates solution paths. In addition, it uses data mining techniques to extract a partial task model from past users' solutions.

Another VEPT that integrates an automatic tutor and was evaluated with an empirical study is TRANSoM (Pioch, Roberts, & Zeltzer, 1997), which is a virtual environment for training pilots of remotely operated underwater vehicles. It was evaluated by Fletcher (1998) with a training transfer experiment. This experiment analyzed the learning effectiveness of the virtual environment for maneuvering skills using a remotely operated underwater vehicle by means of pre- and posttests.

The review of the state of the art shows that there are just a few empirical studies that involve VEPTs. Only a few VEPTs were integrated with an ITS, which is expected to contribute an educational added value by emulating a human tutor in some tasks, but unfortunately, most of the VEPTs equipped with an ITS have not reported empirical evaluations with students that provide evidence on the educational value of this kind of systems. In conclusion, more research is needed to investigate the extent to which this kind of systems are well received by the students and can ease the learning of concepts and procedures in different application domains.

The main purpose of this article is to extend the existing set of empirical results supporting the educational value of virtual environments with a study related to a virtual laboratory for procedural training in biotechnology. Moreover, our work has explored some issues that have been poorly addressed in previous works. Particularly, our virtual lab was used in the context of a first-year course on biotechnology at the university where several students had to learn to perform a long and extensive procedure (around 120 actions) in a virtual world, sharing installations and instruments. In addition, the students, after receiving a short training in using virtual worlds, had to work at home, using his own computer, and only with the guidance of the automatic tutor. One of our goals was to investigate whether this demanding context of use would be too high a barrier to allow students to learn and enjoy the virtual practice.

Description of the Virtual Laboratory

We have implemented the VLB relying on the OS platform. OS supports the development of virtual worlds where several users can simultaneously connect to a virtual world deployed in a server. To connect to the virtual world, each user needs a client application or viewer, which allows him or her to control his or her avatar (his or her alter ego in the virtual world) by using the keyboard and the mouse. In this way, the user can navigate along the virtual world, interact with the world's objects, and chat with other users. There are some client applications available for visiting OS worlds, and among them, we chose Firestorm (Amaral, Avila, & Tarouco, 2012) because it is open source and one of the most popular

and complete (<http://www.hypergridbusiness.com/2017/01/opensim-viewer-survey/>). The features of the viewer are important not only for the users but also for the developers. For example, Firestorm provides better content export or import facilities and performs a better memory management than other viewers. After some initial tests with professors in biotechnology, we decided to adapt the graphical user interface (GUI) of Firestorm in order to simplify it, by removing those elements which were unneeded to interact with VLB, and by adding a new button to drop objects more easily than with the default GUI. These minor changes helped to enhance the usability of Firestorm when using it with VLB.

VLB supports several students doing a virtual biotechnology practice (see Figure 1) in which they have to learn the function of a gene encoding a protein responsible for protecting poplar trees against diseases. To do this, each student has to rapidly produce and genetically transform a large number of poplar plant progenies in the virtual laboratory to study the function of this protein. Achieving this requires following specific procedures and using some materials typically available in a biotechnology laboratory, including equipment such as laminar flow cabinets, beakers, tubing, ice trays, ovens, autoclaves, and of course, growing plant material. Typically, undergraduate students do not have access to real biotechnology labs with this kind of material, and even if they had, this practice would have a cost of 5,000 to 25,000 euros per student in real life. This is unaffordable for a public higher education institution. Moreover,

Figure 1. Snapshot of the virtual laboratory of biotechnology.

the whole task practice would typically take 2 years in real life. This is much longer than the length of a course. Our virtual laboratory overcomes these two problems by speeding up the experiment as much as necessary and by removing any costs derived from the use of expensive laboratory materials.

The task practice is designed as a closed-ended problem, in the sense that there are not many ways to attain the final goal of the practice. Students have to follow a preestablished protocol, although, in some phases of the practice, they are free to perform some actions in different orders, for example, when they have to collect a set of chemicals from different places.

The learning activity combines highly structured parts and less-structured parts, carefully selected by the professor who designed the practice. In the highly structured parts, the student is told of which action he or she should take next, and he or she is only allowed to make irrelevant errors. On the other hand, in the less-structured parts of the practice, the student is given only a general hint instead of being told specifically of which action to take next. For example, when the student has to choose the chemicals to be added to a mixture, he or she is instructed to prepare a mixture but not which specific chemicals are required. This design decision is consistent with productive failure (Kapur, 2008) and we have observed that it results in improved learning.

The design of the virtual practice borrows some features from computer games: The practice to be performed is divided into phases or challenges, each of which must be overcome in order to move forward to the next phase. At the end of each phase, the student is informed of his or her mistakes, if any. However, for pedagogical reasons, even if the student makes mistakes, the automatic tutor can allow him or her to pass to the next phase of the practice. The activity of the students in the virtual laboratory is logged for later examination by the professor or the development team.

VLB is equipped with an automatic tutor that is able to validate the students' actions and to provide assistance to students during the practice, by showing them their mistakes or by giving them some indications about the next actions to be performed in the practice, whenever it is pedagogically appropriate.

Some videos of students using VLB can be accessed at <http://3dlabs.upm.es/laboratorios.php#lab16>.

Description of the Automatic Tutor

Our automatic tutor validates the user attempts at actions and provides them with 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 (see top left snapshot in Figure 2). Big balloons are zoomed images of the error messages warning students about wrong actions or objects required in the next action of the practice. This type of messages is also used to report, at the end of each phase, any errors that the student has made so far

Figure 2. Examples of automatic tutor's assistance.

(lower snapshot in Figure 2). The type of assistance that will be given to the student at any time is configured in a “tutor configuration” file whose content is easily understandable and modifiable by a teacher without programming skills. The syntax of this file was described by Rico, Ramírez, Riofrío, Berrocal-Lobo, and De Antonio (2012).

In OS, objects communicate to each other through the exchange of messages. Figure 2 shows the sequence of messages that the objects involved in a fragment of a practice exchange among them while the student is performing the practice. In this fragment, the student has to prepare a mix of three chemicals in a tube. These chemicals are kept in a refrigerator, so first the student has to take an ice tray (Step 1). Then he or she has to take these chemicals from the refrigerator (Steps 2 and 3), carry them on the ice tray to the workspace, and drop the ice tray on the table (Step 4) where the mixture will be prepared. Next, before adding each chemical to the mix (Step 11), the student needs to get a micropipette with a tip (Steps 6, 7, and 8). Finally, the student takes the tube with the mix (Step 12) and carries it to another place of the lab.

For the sake of clarity, the validation request for each action attempt 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 exchanged for each

user action. If the action is right, the tutor replies OK, otherwise the tutor can reply in two different ways depending on the tutoring strategy configured: If the reply is FAIL, then the tutor denies the permission to perform the action and the student is informed of the reason; otherwise, if the reply is FAIL BUT GO, the tutor gives the permission to perform the action, but it registers the mistake without warning the student immediately. In this way, students can discover later on the consequences of their mistakes because these consequences are also simulated in the virtual world (in our example, the final result would be a dead plant).

Pedagogic Evaluation

Method

We were interested in studying the educational effectiveness of VLB. To this end, we designed an experiment in which we aimed to evaluate to what extent VLB helps to learn new concepts and procedures or to consolidate them. In addition, we also pursued to measure the degree of usability, the perceived educational value, and the satisfaction of students concerning VLB.

We were interested in testing two main hypotheses:

1. Students that perform the virtual practice acquire concepts that are necessary to complete the practice, even if they haven been explained in previous lectures.
2. Students that perform the virtual practice learn better the concepts previously explained in lectures and applied in this virtual practice, when compared with students that just attend lectures.

To test these hypotheses, we worked with two groups of students: a test group, which performed the virtual practice in VLB; and a control group, which only received conventional training. After receiving conventional lectures and performing the virtual practice (or not), both groups did an exam that allowed us to measure the degree of learning achieved for each group. In addition, we were interested in exploring the kinds of knowledge that were better learnt by means of VLB. To this end, we established two different classifications for the questions in the exam:

1. According to the place where the knowledge needed to answer the question was first presented to students, either in a previous lecture or in the practice itself through some slides or automatic tutoring messages.
2. According to the type of knowledge involved in the questions: environmental condition, chemical substances, and so forth.

Procedure

The experiment was run twice, in the second semesters of 2013 and 2014, in the last month of each semester. Both experiments were conducted under the same conditions. Students performing the practice obtained a grade that had a weight of 5% in the final grade of the subject.

At the beginning of the study, all students received a lecture of 2 h 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, what did not prevent students from succeeding in the practice because all missing information was available within the VLB.

Afterwards, one of the researchers gave students in the test group a talk of 1 h on the procedure of the study and the basics of the avatar's control, and put at their disposal a help card on the avatar's control. Then, students had to install the client application in their own computers and connect to the virtual lab to complete a simple tutorial on their own. In this tutorial they simply had to take a cup of coffee from a vending machine in the virtual world, sit down on a chair, and drink it. The protocol of this tutorial was designed so that students had to employ actions similar to those that they would have to perform in the virtual practice. Once they completed the tutorial, they had to choose a time slot in a specifically created web application to perform the practice. The maximum number of students allowed at the same time slot in the VLB was 6.

The virtual practice comprised around 120 basic actions grouped into three phases. The students needed around 90 min on average to complete the entire practice. Throughout the practice, each student was assisted by the automatic tutor, as explained before in the description of the virtual laboratory. We wanted students to use their own computers and Internet connections to perform the practice without the physical presence of other students, technical staff, or a professor; however, an avatar controlled by a member of the technical staff was always in the virtual lab to help the students via chat with any issue related to the avatar's control or the location of some objects. 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 protocol of the practice than the one provided in the slides presented in the initial lecture.

As soon as the students completed the whole practice, they had to fill in an impressions questionnaire about the virtual laboratory.

As a great number of students enrolled for the virtual practice, and in order to have a similar number of students in the test group and in the control group, we decided to conduct the exam on the biotechnology concepts before all the students enrolled had performed the practice. The students were not warned of this exam in advance. In this way, the knowledge of two different groups was compared: The group that had already performed the practice (test group) and the

group of students that were still waiting for their time slot or were not going to do the practice (control group). Before doing the exam, and to provide an alternative way of learning the same concepts as the students of the test group, students of the control group watched a video of an avatar performing the entire practice in the virtual lab; 27 students performed the exam in 2013 and 39 in 2014, where 15 and 23, respectively, had actually performed the practice (test group); and 12 and 17, respectively, had not (control group).

Because the composition of the groups was the result of the time slots chosen by the students and their willingness or not to perform the virtual practice, a pretest was carried out to verify the homogeneity of the groups, that is, to check whether the test group and the control group were formed by students with a similar academic level. The pretest grade was obtained from a previous exam of the same subject. Results are described later.

Measures—Instruments

As mentioned in the previous section, the students performed a surprise exam on biotechnology concepts. The exam consisted of 17 multiple choice questions with four or five available choices per question and just one right answer. Aiming at controlling the effect of random guessing, a wrong answer had a penalty of $X/(\text{number of choices} - 1)$ where the right answer scored X points, and a blank answer scored 0 points.

To better assess the kind of impact of the practice in the students' knowledge, three types of questions, considering the source of the knowledge required to answer, were included in the exam (in parenthesis the number of questions in each group):

- A1. Questions related to concepts that could only be acquired by performing the practice (or watching the avatar performing the practice in the video) (7),
- A2. Questions related to concepts that were explained in the lecture but were not needed in the practice (1), and
- A3. Questions related to concepts that were explained in the lecture and had to be applied in the practice (9).

Moreover, the questions were classified from a second perspective, by the type of knowledge required to answer each question. In this classification, four types of questions were considered:

- B1. Environmental condition required in some step of the practice, such as temperature, relative position, and so forth (3);
- B2. Chemical substances required in some step of the practice (4);
- B3. Purpose of a machine required in some step of the practice (3); and
- B4. Instruments required in some step of the practice (2).

The rationale of this second classification is to analyze what types of knowledge were more easily learnt or consolidated in the virtual lab.

At the end of the practice, a questionnaire was presented to each participant. This questionnaire included three open questions where students had to enumerate positive and negative aspects of the lab and to express additional comments on the lab, respectively.

Results

Impact in Learning

For the analysis of the homogeneity of the test and control groups, we performed a Student's T test with $\alpha=0.05$ for checking the equality of means on the pretest grades. The result was that the means in both groups were equivalent (sig [bilateral]=0.175 in 2013 and sig [bilateral]=0.179 in 2014). Hence, both groups were not significantly different.

In addition, we also checked if the grades obtained in the surprise exam and in each subset of questions followed a normal distribution in each group. To this end, we employed a Shapiro–Wilk test with $\alpha=0.05$. The result was that the global grades and the grades of the Subsets A1 and A3 followed a normal distribution in both groups. However, we could not demonstrate that Subset A2 followed a normal distribution. Despite of it, we decided to include in this section the results related to Subset A2 because we think that they also provide valuable information.

Finally, we checked whether the dependent variables (measured by average grades in A1 questions and average grades in A3 questions) depend on the independent variable (having performed the virtual practice or not) by applying a Student's T test with $\alpha=0.05$. The result indicated that there was indeed dependency, with the following statistical significance.

The differences between the average grades obtained by both groups (expressed as positive or negative percentages) in the subsets of questions derived from the first classification (source of the knowledge), and globally, are summarized in the following table.

The students that performed the virtual practice (test group) obtained grades in the exam that were better (22.09% and 20.21%) than those of the students in the control group. If we only focus on questions of Type A1 or A3, the grades of the test group were also better than the ones achieved by the control group, with a greater difference in A1 questions (related to new concepts that are necessary to complete the practice but are not explained in lectures).

The average grades (on a 0–10 scale) obtained by the test group in the subsets of questions derived from the second classification (type of knowledge) are summarized in the following table.

Here, it is worth mentioning that in both experiments, questions of Type B1 led to the lowest grades.

Discussion

Impact in Learning

The results obtained seem to support both Hypothesis 1 and 2. Concerning Hypothesis 1, as questions of Type A1 are associated with new concepts only presented in the virtual practice, results reveal that students in the test group, who really performed the practice, as expected, learnt these concepts better than students in the control group, who only watched an avatar performing the practice. The average grades in A1 questions were significantly better for the test group in both years (see Table 1 for statistical significance of the results and Table 2 for the magnitude of the difference).

Regarding Hypothesis 2, as questions of Type A3 are associated with concepts explained in the initial lecture that were also applied in the virtual practice, results indicate that students in the test group, also as expected, consolidated these concepts better than students in the control group. The average grades in A3 questions were significantly better for the test group in both years (see Table 1 for statistical significance of the results and Table 2 for the magnitude of the difference).

Table 1. Study of Dependency Between Variables.

	Type A1 (only practice) Sig (bilateral)	Type A3 (only lecture) Sig (bilateral)
Semester 2013	0.009	0.005
Semester 2014	0.005	0.048

Table 2. Difference Between the Grades in the Two Groups.

	Type A1 (only practice) grade(Test) – grade(Control)	Type A2 (only lecture) grade(Test) – grade(Control)	Type A3 (practice + lecture) grade(Test) – grade(Control)	All grade(Test) – grade(Control)
Semester 2013	29.43%	–35.65%	19.84%	22.09%
Semester 2014	20.46%	9.52%	15.81%	20.21%

Table 3. Grades Obtained by the Test Group According to the Second Classification.

	Type B1 (condition)	Type B2 (substances)	Type B3 (purpose)	Type B4 (instrumental)
Semester 2013	3.2	7.11	4.11	5.5
Semester 2014	1.8	4.57	5.07	4.06

On the other hand, the execution of the virtual practice does not seem to have any impact on questions of Type A2, as different results are obtained in each year.

In Table 3, we can find the results from the perspective of the second classification (type of knowledge). Here, it is worth pointing out that the grades in the exam were generally low, probably because, as mentioned in the measures and instruments section, students had to take the exam by surprise. In our opinion, the finding that the lowest grade is obtained in questions of Type B1 (required environmental conditions) can be explained by the fact that these concepts are not graphically represented in any way in the virtual world, but instead they have to be learnt by heart, like the required temperature for some process. If we analyze the results obtained in questions of Type B2 to B4, we cannot conclude that the students of the test group learn better some type of knowledge than others. Thus, more experiments will be necessary to further explore the impact of the virtual practices in the learning of the different types of knowledge.

Another point that deserves a clarification is why the results of 2014 are slightly worse than the results of 2013. We believe that this difference can be explained because, in 2014, the exam on the virtual practice was conducted closer to the final exam date. Therefore, the difference of knowledge between the test group and the control group could have been reduced by the increased study time in both groups, especially evident in questions of Type A3 (see high sig bilateral in Table 1).

Although the impact of the virtual practice on learning seems to be reasonably good in these two cohorts of students, we believe that the results might have been even better if postgraduate students had been involved. We base our belief in that postgraduate students generally find more motivating and stimulating the possibility to have access to virtual practices than first-year students because postgraduate students are indeed more interested in biotechnology and understand better the benefits that VLB can offer them. Furthermore, we hypothesize that if students had the chance of doing more virtual practices in the future, they would become more familiar with this way of learning and the technology would become even more effective. These beliefs pose new hypotheses that future research should explore.

One of the main limitations in our study is the size of the exam, which was intentionally limited by the instructors of the subject. The more questions in the

exam, the more reliable would be the results. However, at this point, we should clarify that the questions of Type A3 in the exam covered very well the concepts explained in the classroom that are also applied in the virtual practice. However, it would be desirable to extend the other two sets of questions.

Qualitative Analysis of Impressions

In this section, we will present the qualitative analysis of the three open questions filled out by students (positive aspects, negative aspects, and additional comments). Neither of these questions was mandatory, but most of the students answered them. Frequently, the additional comments field was filled out with positive and negative impressions.

For the analysis of the answers, we carried out a category-construction process (Merriam, 1998) that yielded a first level of basic categories directly derived from students’ answers. Then, we grouped together these basic categories into different general categories, which resulted in an intermediary level of categories. Finally, we grouped together the categories of this intermediary level into more general categories in a top level.

Concerning negative aspects, we derived 47 basic categories that yielded 7 categories for the intermediary level and 3 categories for the top level (as shown in Figure 3). Next, we will summarize the most relevant students’ opinions behind each of the seven categories of the intermediary level:

LABLIM: It encompasses comments related to limitations of the lab. For example, some students observed some bugs in animations or perceived a slow feedback of

Figure 3. Taxonomy of categories on negative aspects.

the system in some moments of the practice. A few students complained that the reasons of a student's mistake are not always well explained.

DIFAVATARHANDLE: These comments are related to difficulties with the handling of the avatar and the chat interface. Some students highlighted their difficulties to control the avatar through the keyboard and the mouse, more precisely a few students mentioned the pickup or drop operations and the management of the inventory. One student complained that tutor's messages were mixed with students' messages in the same window chat, and another one commented that the GUI of the viewer creates a false illusion of complexity because it contains many elements that are not necessary to do the practice.

PRACTICEDESIGN: It comprises answers in which students showed their dissatisfaction with some features of the practice that have been designed like that on purpose. A significant number of students would have liked to have something like an "undo command" that allowed them to undo an incorrect action when they realized the mistake. Even more, one student pointed out that this lack is frustrating. This command was not included because we aimed (supported by the professor of the subject) at implementing a realistic practice where students have to take care of not making mistakes, like in the real world. However, even if students make serious mistakes in a phase of the practice, they are allowed to pass to the next phase and begin from a valid state (as if they had not made any mistake). Another feature that students missed was the possibility of performing actions out of the protocol of the practice, so that students had more freedom to explore and learn by themselves other concepts in the lab. Finally, one student mentioned that it would be very interesting to be able to perform the practice along a period in different days instead of a prefixed slot in a day.

BADPERCEIVEDUTIL: This category comprises the comments of a few students who got a bad impression of the educational value of the lab; for example, some said that the practice had not served them to learn too much.

UNREALEASE: In these comments, students complained about aspects of the lab that paradoxically are the same as or even better than in a real lab. Here, we can find comments in which some students complained that the practice was too long or that they had to go to a special place in the lab every time they needed to see the details of the protocol. Moreover, a few students complained that they had to visit more than one room, sometimes far from each other, to do the practice, or that they had to wait for some instruments while they were being used by other students. Therefore, it seems they would expect things in the virtual lab to be easier than in the real world.

DIFPRACTICE: It comprises comments where they complained of some features of the practice or lab that make the practice more difficult. For example,

a significant number of students complained that some steps are not well indicated because in some moments, the automatic tutor does not give enough details on the next actions to be done or how to do them. However, this was decided on purpose for pedagogical reasons. In addition, other students highlighted the difficulty of finding some of the required materials to do the practice, which can be explained because students are not familiar with the appearance of many lab objects. A few students said that it is very difficult to perform the practice without studying a lot the subject matter before.

CONTENTERROR: These comments refer to minor inconsistencies between the terms studied in the classroom and the terms that appear in the lab or in the tutor directions.

In the top level of the hierarchy, as shown in Figure 3, we have the following categories: **LABISSUE**, which encompasses categories related to the design of the virtual lab; **DOMAINCONSTR**, which comprises categories related to constraints derived from the application domain or the real practice; and **CONTENTISSUE**, which represents issues associated with the learning material existing in the virtual lab.

One of the most evident conclusions we have reached from the negative comments is that many students would like to have more help and features in the practice (see **DIFPRACTICE-**, **PRACTICEDESIGN-**, and **UNREALEASE-** related comments). Here it is worth noting that this request may be conflicting with constructivist pedagogical strategies, where students are forced to solve problems in which their current level of knowledge is challenged. A virtual lab with much more help would make students feel more comfortable but it would lead to worse learning results. In addition, we think that the comments behind the **UNREALEASE** category reveal that some students are not aware of how a real lab or the real practice are. It is also interesting to note the request of an “undo command,” probably because they are used to find this command in other software applications, where it may be acceptable because they are not simulating the real world.

Concerning positive aspects, as shown in Figure 4, we derived 29 basic categories in the bottom level, 7 categories in the intermediary level, and 2 categories in the top level. Next, we will summarize the most relevant students’ opinions behind each of the seven categories of the intermediary level:

EDUVALUE: It encompasses comments where students highlighted the educational value of the lab. Here, many students said that this practice represents a good way of learning. A significant number of students said that this lab allowed them to see how a real lab is and another significant group said that this practice was a fruitful experience, but it might have been even better with the suitable resources (this same group of students complained of some performance issues,

Figure 4. Taxonomy of categories on positive aspects.

see LABLIM category). Some students said that the experience was motivating and others said that they would like to repeat the practice to consolidate the concepts.

VRADVANT: The comments of this category reveal that a significant number of students realized the advantages of a virtual lab with respect to a real lab. Some of them said that in the virtual lab, they could carry out the practice more quickly. Other students appreciated that they were able to perform a practice that in the real world requires resources inaccessible to them. A few of the students highlighted that they could perform the practice without risks and in a more comfortable fashion than in the real world.

REAL: In this category, many students' answers praised the quality of the graphics and the fidelity of the lab to reality.

GOODGUIDE: It comprises many comments where students praised the assistance of the automatic tutor and the avatar controlled by a member of the technical staff.

FUN: This category encompasses many comments in which students remarked that the practice was amusing. One student said that handling the avatar was also fun.

GOODHELP: Here a significant number of students praised the quality of the slides and a few of them appreciated the signals existing in the lab such as the directions to the rooms, the labels under the objects, and so forth.

EASYAVATAR: It comprises a couple of comments where students highlighted the ease to handle the avatar.

As shown in Figure 4, the top-level categories in the hierarchy of positive issues are as follows: **GOODPERCEIVEDUTIL** where students have demonstrated having perceived the utility of the virtual lab and **LABQUALITY**, which comprises five categories in which students have highlighted the quality of the virtual lab.

When considering positive comments, the comments behind the **VRADVANT** are especially remarkable because they show that many students, despite their little experience with virtual labs, were able to perceive the benefits of this kind of systems. Moreover, the comments in **GOODGUIDE** related to the assistance of the human avatar are also noteworthy, since some of them can also be seen as limitations of the automatic tutor, mainly regarding its inability to answer students' questions or to encourage students when they look frustrated.

Conclusions

We have presented the results of a study where undergraduate students have employed a cost-effective VWPT, called VLB, as supplemental material to a biotechnology course, and consequently, they learnt more effectively than other students who did not use this tool. Moreover, most of the students that performed the virtual practice have shown their satisfaction with the experience and have evaluated positively the educational value of the tool. Hence, we have provided evidence that this virtual practice is a valid supplemental material for a classical course on biotechnology, which can enhance learning effectiveness. This opens up new possibilities for students that are typically not available in most of the universities mainly because of economic reasons.

On the other hand, we must be cautious with the results due to the small number of evaluated students. Therefore, similar experiments will have to be conducted in the future to back up the conclusions of this study.

From a more general perspective, and given that VLB is a good representative of VWPTs, we conclude that this work contributes new evidence to support the thesis that VWPTs are a helpful tool for education.

Concerning future work, the results of this experiment pose some challenging lines of research. One of them is the study of long-term retention in comparison with conventional teaching methods. We aim to confirm that a tool designed according to the "learning by doing" approach provides better long-term retention than other teaching methods (Dale, 1969). Furthermore, we are also interested in conducting similar experiments in other domains different to biotechnology where we can reuse the same automatic tutor or an improved version. Regarding enhancements to the automatic tutor, currently we are researching how to leverage students' logs to provide a better tutoring feedback. Finally, students' logs may also be processed to build effective visualizations of the group performance to make it easier for teachers to improve the lab and their own teaching.

Acknowledgments

The authors would like to thank UPM Educational Innovation Program (Projects IE 100210147, PT 07017-1112 and IE 12-1307001) and to GIE-LVBA members. Riofrío-Luzcando thanks Secretariat of Higher Education, Science, Technology, and Innovation from Ecuador (SENESCYT), and Rico thanks projects TIN2011-24139 and JCI-201.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

- Amaral, E., Avila, B. G., & Tarouco, L. M. R. (2012). Aspectos teóricos e práticos da implantação de um laboratório virtual no OpenSim [Theoretical and practical aspects of the implementation of a virtual laboratory in OpenSim]. *Anais do Simpósio Brasileiro de Informática na Educação [Proceedings of the Brazilian Symposium on Informatics in Education]*. (Vol. 23), Brazil.
- Buche, C., Querrec, R., & De Loor, P. (2003). MASCARET: Pedagogical multi-agents systems for virtual environment for training. *Proceedings of the International Conference on Cyberworlds*, USA, 423.
- Clark, R. C., & Mayer, R. E. (2011). Simulations and games in e-learning. In *Simulations and Games in e-Learning. In e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning* (pp. 368–398). 3rd Edition, Pfeiffer, San Francisco, CA, USA. DOI: 10.1002/9781118255971.ch16.
- Dale, E. (1969). *Audiovisual methods in teaching* (3rd ed.). Oak Brook, IL: The Dryden Press.
- Dalgarno, B. (2002). The potential of 3D virtual learning environments: A constructivist analysis. *Electronic Journal of Instructional Science and Technology*, 1–19. Retrieved from <http://doi.org/Journal-article>
- Dalgarno, B., & Lee, M. J. W. (2012). Exploring the relationship between afforded learning tasks and learning benefits in 3D virtual learning environments. In M. Brown, M. Hartnett, & T. Stewart (Eds.), *Proceedings of ASCILITE-Australian society for computers in learning in tertiary education annual conference* (pp. 236–245). Wellington, New Zealand: Australasian Society for Computers in Learning in Tertiary Education.
- Fletcher, B. (1998). TRANSOM ROV pilot training: Test results. *Underwater Intervention '98 Conference Proceedings*, Washington, DC.
- Fournier-Viger, P., Nkambou, R., Nguifo, E. M., Mayers, A., & Faghihi, U. (2013). A multiparadigm intelligent tutoring system for robotic arm training. *IEEE Transactions on Learning Technologies*, 6(4), 364–377.
- Gu, Y., Sosnovsky, S., & Ullrich, C. (2015). Safechild: An intelligent virtual reality environment for training pedestrian safety skills. In G. Conole, T. Klobučar, C. Rensing, J.

- Konert, & É. Lavoué (Eds.), *Design for teaching and learning in a networked world SE – 11* (Vol. 9307, pp. 141–154). Springer International Publishing. Retrieved from http://doi.org/10.1007/978-3-319-24258-3_11
- Holzinger, A., Kickmeier-Rust, M. D., Wassertheurer, S., & Hessinger, M. (2009). Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the haemodynamics simulator. *Computers & Education*, 52(2), 292–301. Retrieved from <http://doi.org/10.1016/j.compedu.2008.08.008>
- Hospers, M., Kroezen, E., Nijholt, A., op den Akker, R., & Heylen, D. (2003). An agent-based intelligent tutoring system for nurse education. In A. Moreno & J. Nealon (Eds.), *Applications of software agent technology in the health care domain SE – 10* (pp. 143–159). Birkhäuser Basel. Retrieved from http://doi.org/10.1007/978-3-0348-7976-7_10
- Huang, H.-M., Rauch, U., & Liaw, S.-S. (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education*, 55(3), 1171–1182.
- İnce, E., Kırbaşlar, F. G., Güneş, Z. Ö., Yaman, Y., Yolcu, Ö.Yolcu, E. (2015). An innovative approach in virtual laboratory education: The case of "IUVIRLAB" and relationships between communication skills with the usage of IUVIRLAB. *Procedia – Social and Behavioral Sciences*, 195, 1768–1777. Retrieved from <http://doi.org/10.1016/j.sbspro.2015.06.377>
- İnce, E., Kırbaşlar, F. G., Yolcu, E., Aslan, A. E., Kayacan, Z. C., Alkan Olsson, J., . . . Charalambis, D. (2014). 3-dimensional and interactive istanbul university virtual laboratory based on active learning methods. *Turkish Online Journal of Educational Technology-TOJET*, 13(1), 1–20.
- Johnson, W. L., & Rickel, J. (1997). Steve: An animated pedagogical agent for procedural training in virtual environments. *SIGART Bulletin*, 8, 1–4.
- Johnson, W. L., Shaw, E., Marshall, A., & LaBore, C. (2003). Evolution of user interaction: The case of agent adele. In *Proceedings of the 8th International Conference on Intelligent User Interfaces* (pp. 93–100). New York, NY: ACM. Retrieved from <http://doi.org/10.1145/604045.604063>
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379–424. Retrieved from <http://doi.org/10.1080/07370000802212669>
- Larsen, C. R., Soerensen, J. L., Grantcharov, T. P., Dalsgaard, T., Schouenborg, L., Ottosen, C., . . . Ottesen, B. S. (2009). Effect of virtual reality training on laparoscopic surgery: Randomised controlled trial. *BMJ*, 338, b1802.
- LeBlanc, V. R., Urbankova, A., & Hadavi, F. (2004). A preliminary study in using virtual reality to train dental students. *Journal of Dental Education*, 68(3), 378–383.
- Los Arcos, J. L., Müller, W., Fuente, O., Orúe, L., Arroyo, E., Leaznibarrutia, I., & Santander, J. (2000). LAHYSTOTRAIN: Integration of virtual environments and ITS for surgery training. In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Intelligent tutoring systems SE – 8* (Vol. 1839, pp. 43–52). Springer Berlin Heidelberg. Retrieved from http://doi.org/10.1007/3-540-45108-0_8
- Merriam, S. B. (1998). *Qualitative research and case study applications in education. Revised and expanded from case study research in education*. San Francisco, CA: Jossey-Bass Publishers.

- Müller-Wittig, W. K., Bisler, A., Bockholt, U., Los Arcos, J. L., Oppelt, P., Stähler, J., & Voss, G. (2000). LAHYSTOTRAIN development and evaluation of a complex training system for hysteroscopy. *Studies in Health Technology and Informatics*, 81, 336–340.
- Pioch, N. J., Roberts, B., & Zeltzer, D. (1997). A virtual environment for learning to pilot remotely operated vehicles. *Proceedings of International Conference on Virtual Systems and MultiMedia VSMM '97*. Retrieved from <http://doi.org/10.1109/VSMM.1997.622350>
- Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N. T., Koedinger, K. R., Junker, B., . . . Choksey, S. (2005). The assistment project: Blending assessment and assisting. *Proceedings of the 12th Annual Conference on Artificial Intelligence in Education*, Amsterdam, 555–562.
- Rico, M., Ramírez, J., Riofrío, D., Berrocal-Lobo, M., & De Antonio, A. (2012). An architecture for virtual labs in engineering education. In *IEEE Global Engineering Education Conference (EDUCON)*, Marrakesh, Morocco, 1–5.
- Ritter, S., Anderson, J., Koedinger, K., & Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, 14(2), 249–255. Retrieved from <http://doi.org/10.3758/BF03194060>
- Ritter, S., Carlson, R., Sandbothe, M., & Fancsali, S. E. (2015). Carnegie learning's adaptive learning products. *Educational Data Mining 2015: 8th International Conference on Educational Data Mining*, Madrid, Spain.
- Shaw, E., Johnson, W. L., & Ganeshan, R. (1999). Pedagogical agents on the web. In *Proceedings of the Third Annual Conference on Autonomous Agents* (pp. 283–290). New York, NY: ACM. Retrieved from <http://doi.org/10.1145/301136.301210>
- Sleeman, D., & Brown, J. S. (1982). *Intelligent tutoring systems*. London, England: Academic Press.
- Song, H. S., Pusic, M., Nick, M. W., Sarpel, U., Plass, J. L., Kalet, A. L. (2014). The cognitive impact of interactive design features for learning complex materials in medical education. *Computers & Education*, 71, 198–205. Retrieved from <http://doi.org/10.1016/j.compedu.2013.09.017>
- Zhong, Y. (2013). *iVirtualworld: A domain-oriented end-user development environment for building 3D virtual chemistry experiments*. Athens, OH: Ohio University.
- Zhong, Y., & Liu, C. (2012). User evaluation of a domain-oriented end-user design environment for building 3D virtual chemistry experiments. *User Evaluation for Software Engineering Researchers (USER)*, 5–8. Retrieved from <http://doi.org/10.1109/USER.2012.6226585>.