International Journal of Human-Computer Interaction

Evaluation of internal and external validity of a virtual environment for learning a long procedure

--Manuscript Draft--

Manuscript Number:	IJHC-D-16-00246R2
Full Title:	Evaluation of internal and external validity of a virtual environment for learning a long procedure
Article Type:	Original Research
Section/Category:	Foundations of Cognitive Science for the Design of HCI Systems
Corresponding Author:	Charlotte Hoareau, Ph.D. Universite de Bretagne Occidentale Plouzané, FRANCE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universite de Bretagne Occidentale
Corresponding Author's Secondary Institution:	
First Author:	Charlotte Hoareau, Ph.D.
First Author Secondary Information:	
Order of Authors:	Charlotte Hoareau, Ph.D.
	Ronan Querrec, Associate Pr.
	Cédric Buche, Associate Pr.
	Franck Ganier, Pr.
Order of Authors Secondary Information:	
Abstract:	Virtual reality is a frequently used tool in vocational training. Nevertheless its efficiency has not been systematically tested. The main goal of this study is to assess the effectiveness of a virtual environment (VE) for learning a complex procedure in the biomedical domain. Two experiments were performed. The first one assessed internal validity of the VE, which is the effectiveness of using a VE in the process of learning a new procedure. The second one tested external validity of the VE, which is the participants' ability to reproduce the acquired skills in a real context. We find that internal and external validity must be evaluated before using a virtual environment in long-term procedure learning. The results of such evaluations are a basis for future experiments aiming to optimize learning conditions in a VE and transferring the acquired skills in a real context.

Evaluation of internal and external validity of a virtual environment for learning a long procedure

Charlotte Hoareau[1,2]

Ronan Querrec [2,3]

Cédric Buche [2,3]

Franck Ganier [1,2]

[1] Université de Bretagne Occidentale, F-29200 Brest, France

charlotte.hoareau@univ-brest.fr

franck.ganier@univ-brest.fr

[2] Lab-STICC CNRS UMR 6285

European Center for Virtual Reality

Technopôle Brest-Iroise F-29280 Plouzané, France

[3] Ecole Nationale d'Ingénieurs de Brest, F-29280 Plouzané, France

querrec@enib.fr

buche@enib.fr

Correspondence concerning this article should be addressed to Charlotte Hoareau (charlotte.hoareau@univ-brest.fr).

Author biographies

Charlotte Hoareau obtained her Ph.D. in Cognitive Psychology and Ergonomics at the University of Western Brittany in 2016. She is a member of the IHSEV team of the Lab-STICC laboratory. She is specialized in procedural learning, virtual environment for training and assessment of new interfaces as well as virtual environments.

Ronan Querrec received a habilitation to direct researches degree in 2010 and a PhD degree in 2001. He is a member of the board of director of the French Association for Virtual Reality (AFRV). He is at the head of the CERV (Research Center for Virtual Reality).

Cédric Buche received a habilitation to direct researches degree in 2012 and a PhD degree in 2005. He is a member of the board of director of the French Association for Artificial Intelligence (AFIA). He is the Editor-in-Chief of the International Journal of Virtual Reality (IJRV).

Franck Ganier is a professor of Cognitive Psychology and Ergonomics at the University of Brest, France. He is a member of the Lab-STICC Laboratory (UMR 6285 CNRS), European Centre for Virtual Reality (CERV), Brest, France. His research interests concern procedural learning, processes for comprehending instructions and methods for evaluating technical documents as well as virtual environments.

Evaluation of internal and external validity of a virtual environment for learning a long procedure

Abstract

Virtual reality is a frequently used tool in vocational training. Nevertheless its efficiency has not been systematically tested. The main goal of this study is to assess the effectiveness of a virtual environment (VE) for learning a complex procedure in the biomedical domain. Two experiments were performed. The first one assessed internal validity of the VE, which is the effectiveness of using a VE in the process of learning a new procedure. The second one tested external validity of the VE, which is the participants' ability to reproduce the acquired skills in a real context. We find that internal and external validity must be evaluated before using a virtual environment in long-term procedure learning. The results of such evaluations are a basis for future experiments aiming to optimize learning conditions in a VE and transferring the acquired skills in a real context.

Keywords: Procedural Learning, Transfer of learning, Virtual Environment for Training, Virtual Reality, Assessment, Internal validity, External validity.

EVALUATION OF INTERNAL AND EXTERNAL VALIDITY OF A VIRTUAL ENVIRONMENT FOR LEARNING A LONG PROCEDURE

1. Introduction

In order to reduce the time and costs of training, virtual reality has been developed for use in the industrial, military, medical and sport domains (Beyer-Berjot, Palter, Grantcharov, & Aggarwal, 2014; El-Chaar, Boer, Pedrazzoli, Mazzola, & Maso, 2011; Gallagher, Seymour, Jordan-Black, McGlade, & Satava, 2013; Greunke & Sadagic, 2016; Jou & Wang, 2013; Lam, Sundaraj, & Sulaiman, 2014; Miles, Musembi, Pop, & John, 2013; Paiva, Machado, & Valenca, 2013; Park, Jang, & Chai, 2006; Rauter, Sigrist, Koch, & Crivelli, 2013; Spruit, Band, Hamming, & Ridderinkhof, 2014). A number of studies have focused on evaluating the effectiveness of Virtual Environment for Training (VET ; Annetta, Minogue, Holmes, & Cheng, 2009; Dorn & Barker, 2005; Gallagher & Cates, 2004; Wrzesien & Raya, 2010). Virtual reality is increasingly considered as a technology offering some possibilities for learning (Kozak, Hancock, Arthur, & Chrysler, 1993). For the past fifteen years, many applications have been developed but few of them proposed a validation of its real effects on learning. The effectiveness of virtual environments for training has already been studied (Annetta et al., 2009; Chen et al., 2016; Dede, Salzman, Loftin, & Ash, 1997; Huegel, Celik, Israr, & O'Malley, 2009; Loukas, Nikiteas, Kanakis, & Georgiou, 2011; Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014; Wrzesien & Raya, 2010), but first, their conception and/or assessment do not yet meet the rigorous criteria set by experimental design in ergonomics or cognitive psychology, and second the observed effects are not explicitly related to real cognitive process involved when learning. Thus, methodological bias in some studies and/or their lack of reference to the existing literature in ergonomics or cognitive psychology make it difficult to prove the effectiveness of VE (used for training Girard, Ecalle, & Magnan, 2013). The interest of this technology lies in the assumption that what has been learned in a VE can be transferred to real conditions. In this paper, we focus on the assessment of training on VET: the confirmation of internal and external validity of a VE for training.

The constitutive dimensions of the validation of a VET for procedural learning can be defined by an internal and an external validity. We propose the concept of validity (internal and external) for the assessment of VET. A VET demonstrating a high level of internal validity would be one that would maximize learning. In other words, learning performance would reveal the internal validity of the VET. In the context of procedural learning, internal validity of the VET can be validated through the analysis of learner performance. The confirmation of the acquisition of a procedure can be done through the interpretation of the learning curve formed by the performance of learners (e.g. data relating to the completion time of the procedure, the number of errors, etc. ; Kho, 2011; Kim, Ritter, & Koubek, 2013; Thorndike & Woodworth, 1901). Complementarily, VET demonstrating a high level of external validity would be one that would optimize the transfer of knowledge. VET is considered as useful if the knowledge acquired during the training phase is transferable to a real situation. According to Anastassova et al. (2007), in the area of transfer, transfer quality evaluates the effectiveness of learning (Loup-Escande, Burkhardt, & Richir, 2013). In other words, the transfer performance would reveal the external validity of the VET. In sum, the validity of a VET must be represented as a balance between internal validity (i.e. the assessment of learning in controlled conditions) and external validity (i.e. the evaluation of transfer to a natural situation). Effective VET must show a high internal and external validity: it must maximize learning and ensure its transfer.

We propose to measure the validity of a virtual environment created for learning a complex procedure for running blood tests. By using blood tests as a complex procedure we can evaluate the ability to perform these skills, acquired in a virtual context to a real (external validity). The results of the previous studies dealing with virtual reality, procedural learning, and transfer of learning from a virtual environment to reality will be summarized. We then detail our study, which assess the internal validity of a virtual environment by checking the acquisition of a procedure via a learning curve and, studies its external validity by checking the execution of this procedure in the real world. Our results provide a step towards an ergonomic approach assessing first the validity (both internal and external) of a virtual environment, prior to use it as a tool for training.

1.1. Virtual reality and procedural learning

It is not easy to define Virtual Reality as it deals with a lot of aspects (from technology to content model) or research domains (from Cognitive science to Mechanics). Virtual reality has been first defined by Fuchs (1996) in the 90's from a technical point of view as "a scientific and technical field using information technology and behavioural interfaces in order to simulate in a virtual world the behaviours of 3D entities, interacting in real time with each other and with one or more users in real time, through pseudo-natural immersion via sensory motor channel" (cited by Arnaldi, Fuchs, & Tisseau, 2003, p. 8). If VR was then only used in some specifics domains, nowadays with the development of technologies (coming essentially from mobile phone industries), we can envisage it use for large audiences (leisure, entertainment…). As this term is ambiguous, lots of researchers prefer to use the term of Virtual Environment (Hale & Stanney, 2015).

But, for us, the major point of Brooks definition of VR (Brooks, 1999) is that the user lives an “experience” that can modify their internal state. This allows us to consider VR as an interesting tool for learning. It has been used in training, because-it can help simulate conditions close to reality (Arnaldi et al., 2003; Ganier, Hoareau, & Devillers, 2013; Kozak et al., 1993), but it alsoallows the learner to act in unusual, critical or dangerous situations (Bliss, Tidwell, & Guest, 1997; Carpentier & Lourdeaux, 2013; Grumbach & Klinger, 2007; Jorge et al., 2013). In some sectors, such as medicine or aeronautics, the use of simulation is required to train as well as to assess skills (Gallagher & Cates, 2004).

In this context of training, we focus specifically on procedural learning. A procedure can be described as “a set of operations and/or actions in order to reach a set goal” (Heurley, 1997, p. 127). Completing these different actions, in a set order, results in the full completion of the task. One of the particularities of procedural learning is that it requires practice (i.e. repetition of the set of actions to complete). When training starts, learners do not know the set of actions to complete and/or they are performing them for the first time. They must strongly rely on the instructions given to them to complete the actions. Then, as training goes on, learners develop representations based on declarative knowledge, which they transform into actions. This transformation requires a certain amount of time and has a substantial cognitive cost (Dixon, Harrison, & Taylor, 1993; Ganier, 2004; Heurley & Ganier, 2006; Kieras & Bovair, 1986). When the procedure has been repeated several times, the learner is then able to retrieve it directly from his/her long-term memory; execution of the task has becomes automatic (i.e. without instructions or errors, Anderson, 1982, 1998). These 3 stages correspond to a learning curve characterized by low performance at the beginning of learning which then increases over time, leading to a plateau. This plateau corresponds to the learner stabilizing his/her performance after reaching optimal execution of the procedure. Traditionally, in training conditions, learners gain declarative knowledge through the use of material (slideshows, paper or digital documents) given by the trainers (Cazeaux, Devillers, & Saint-Romas, 2005). However, procedural knowledge is very different from declarative knowledge, which is based on repeating a set of actions (Anderson, 1998). Thanks to the high level of interactivity and the range of motor actions they offer to learners, virtual environments can enable procedural knowledge acquisition, in particular through practice. The effectiveness of virtual reality for procedural learning has already been shown (Boud, Baber, & Steiner, 2000; Bruwaene, Schijven, & Miserez, 2014; Ganier et al., 2013), but these studies only address relatively simple procedures made up of a short set of actions (i.e. completed within less than 10 minutes). It would be useful to consider what happens with longer procedures as the length of a procedure is a key factor. The use of a long procedure would bring the experiment closer to ecological learning conditions.

We focus on a long procedure because procedures in occupational settings most often require a long series of actions to complete.

The interest of using virtual reality to study procedural knowledge acquisition is not limited to the ability to memorize a procedure. It is also useful for assessing the possibility of transferring knowledge gained in a virtual environment to a real situation (Carlson, Peters, Gilbert, Vance, & Luse, 2015; Ganier et al., 2013; Rose et al., 2000).

1.2. Transfer from virtual to real

In general, learning transfer occurs when an individual performance when completing a set task is influenced by his/her experience, former knowledge or previous performance when completing a similar task (Perkins & Salomon, 1992). More than an expansion of knowledge, procedural transfer entails complex cognitive processes (Bossard, Kermarrec, Buche, & Tisseau, 2004; Haskell, 2000; Tardif, 1999). Transferring skills can be considered as a set of complex processes for the learner to perform since new knowledge, intrinsically connected to the learning context, is difficult to access and apply to another context (Bracke, 1998; Bransford & Schwartz, 2001; Lave, 1988; Presseau, 2000; Wenger & Lave, 1991). There seem to be two key features in the concept of learning transfer: (1) the gap between the context of acquisition and the context of knowledge retrieval, that is to say the similarity of both contexts. It is assumed that the bigger the gap between the knowledge acquisition context and the context of use, the more difficult the transfer will be (with a high cognitive cost for the learner) (Bracke, 1998); and (2) previous performance of the learner when completing the task, that is to say his/her degree of practical experience of the task (Anderson, Reder, & Simon, 1996).

The first significant factor of an effective transfer would depend on the quantity of similar elements between the learning task and the new task to perform. The gap between both situations can be expressed in terms of the number of similar elements. Anderson (1982) used the theory of “common elements” described by Thorndike & Woodworth (1901) to define the common elements theory (Singley & Anderson, 1987). According to Anderson, procedural learning occurs when declarative knowledge, stemming from the instructions provided to the learner, is translated into procedural knowledge. Once this knowledge is translated into skills, they are then anchored to production rules such as “IF [condition] THEN [action]” (“IF I’m in such a situation THEN I will carry out this set of actions”). The triggering condition to these production rules is that the context or the field of use must share common elements with the acquisition context (Pennington, Nicolich, & Rahm, 1995). The learner must thus be able to identify the isomorphic degree between both situations to call upon the adapted skills (Raynal & Rieunier, 2012). It is however difficult to assess the degree of “similarity” needed to enable transfer of learning. All the elements similar to both contexts do not have the same significance (Butterfield & Nelson, 1989; Singley & Anderson, 1987). The question of similarity between two contexts is also raised regarding transfer from virtual to real. Virtual environment developers often focus on the visual rendition of the scene in order to comply with virtual/real isomorphism. On top of staying true to reality, other significant criteria must also be taken into account, such as localization of the different elements with which the learner must interact and responsiveness of the virtual environment (Dalgarno & Lee, 2010; Masciocchi, Dark, & Parkhurst, 2007; Witmer & Singer, 1998). For example, Witmer & Singer (1998) showed that in the case of a driving simulator, appearance and handling of the virtual car should be similar to a real car so that the skills acquired through the virtual environment could be transferred to a real situation.

The second significant factor for effective transfer is the level of practice the learner reached in performing the task to learn. According to Anderson (1982, 1998, 2013), the repeated execution of the procedure, is critical to its automation and thus its acquisition. The transfer of skills from learning context to the context of use is strongly correlated to the time spent on training by the learner. In other words, the ease with which the procedure is retrieved from memory will impact its transfer (Bransford & Schwartz, 2001; Hasselbring & Sherwood, 1987; McKeown & Beck, 1983; Pennington et al., 1995). The impossibility to transfer is attributed to an insufficient basic training by several authors (Klahr & Nigam, 2004; Lee, 1998; Lee & Pennington, 1993; Littlefield, Delclos, Bransford, Clayton, & Franks, 1989). To ensure that learners have sufficient training in the context of procedural learning in a virtual environment, it is possible to determine the number of attempts needed for the learner to master the procedure (Ganier et al., 2013). The learners' performance is thus most often used as an indicator. If the characteristics of the learner's former training (duration, number of repetition of the procedure, etc.) is not taken into account, the conclusions on the effectiveness of transfer from virtual to real can be biased (Bransford & Schwartz, 2001; Pennington et al., 1995).

Homogeneity in the works regarding transfer to a real situation remains difficult to find. Indeed, even if the method used for assessing transfer is often similar, namely by comparing performance of learners taught in a real environment, with others taught in a virtual environment and a control group, the results vary (Carlson et al., 2015; Ganier, Hoareau, & Tisseau, 2014; Hoareau et al., 2013; Oren, Carlson, Gilbert, & Vance, 2012; Rose et al., 1998). Comparing performance of participants trained in a virtual environment to that of participants of a control group (i.e. without training) can help validate the hypothesis of transfer of skills acquired in virtual environment to real conditions. Acquiring knowledge and putting it into practice in a real situation would be proven by higher performance of participants taught in virtual environments when performing the task in situation.

Comparing the performance of the two groups (virtual vs traditional training) would give qualitative indications on learning and transfer from virtual to real. On top of the ability to acquire skills in a virtual environment and being able to perform them in a real situation, by comparing virtual with traditional training the differences (negative and/or positive) of training in a virtual environment could be observed. For example, Kozak et al., (1993) carried out a study which sought to compare the effectiveness of training in a virtual environment in performing a perceptual-motor task. The authors compared three groups of participants: one group trained in a real environment, one group trained in a virtual environment and a control group with no training. Participants had to grasp cans of different colours and place them on corresponding targets. During transfer to real situation, the participants who were trained in a real environment performed better than those who were trained in a virtual one. The authors concluded that for this type of task, learning was specific to the training context. However, Rose et al., (1998), using the same method of evaluation (comparing three groups with different training), as well as a perceptual-motor task in which a ring had to be moved along a bending electric wire, showed no difference between both types of training (real and virtual) in terms of performance during transfer. The training modalities (real or virtual) had no impact on the execution of the procedure after training. These two studies illustrate the difficulty of rigorously comparing different works regarding transfer from virtual to real, in the case of training for a simple perceptual-motor task which calls for sensorimotor abilities rather than acquiring knowledge. Moskaliuk, Bertram, et Cress (2013) underlined the lack of studies concerning the effectiveness of virtual environments for complex tasks. Their study focused on the assessment of a virtual environment developed for complex problem resolution and decision-making tasks involving communication and teamwork for police officers. The results showed that, even though all participants were equivalent regarding knowledge acquisition, the benefits of training through a virtual environment were made clear during transfer. T, participants taught in a virtual environment solved problems more quickly than those taught traditionally. However, the authors stressed that comparing both types of training was difficult as the participants who had been trained with the virtual environment had access to declarative knowledge, provided by the virtual environment, and had been trained on a higher number of scenarios than those who had trained traditionally. In learning a complex procedure, a virtual environment for learning must enable acquisition of knowledge related to this procedure and its transfer to a real situation. During transfer, equivalent performance of learners taught traditionally in a real context and those taught in a virtual environment would be a clear indicator of the effectiveness of virtual environments for learning complex procedure (Anastassova et al., 2007). This is the general objective of our study.

1.3. Objective, materials and scope

The objective of this study is to assess the effectiveness of a virtual environment in procedural learning. The effectiveness will be assessed through procedural learning (experiment 1, section 2) and then its transfer to a real situation (experiment 2, section 3). Thus, after checking that learning a procedure through a virtual environment was possible (internal validity), two types of training were compared (traditional training and training in a virtual environment) to a control situation (external validity).

Both experiments have been conducted in the context of the usage of an automata (STAR) for blood analysis. The procedure used for the experiment (about 125 actions for 13 minutes' duration) has been defined in collaboration with teachers involved in the real educational program. This mean that learning objectives and task analysis have been already done by those teachers. The learners involved in this program are aged 20 to 24 years.

As virtual reality may involve lots of technologies (CAVE, HMD, etc.) or behaviors (moving, grasping, etc.) we will first use a ”Desktop VR” (3D environment, classical screen and mouse interaction) environment in order to isolate all those VR aspects, control the experiments variables and focus on cognitive process involved in learning. Impact of level of presence (immersion/interaction) will be addressed in future works.

2. Experiment 1: Learning a procedure through a virtual environment

The basic idea for this experiment stems from a previous study (Ganier et al., 2013) which showed that a virtual environment could be an effective tool in learning a procedure. The authors used a maintenance procedure for the military Leclerc tank, made up of 25 actions to perform, using written instructions. The results illustrated that after 10 consecutive repetitions of this relatively short procedure, participants were able to reproduce it again a few days later, without referring to the instructions or making errors. The objective of the experiment presented here is to go further by assessing the internal validity of a virtual environment for learning a longer procedure (made up of 125 actions to perform) with audio instructions used as a guide. Learning tasks and recall tasks become more complex because of the length of the procedure. Using a dual-channel presentation: (1) auditory for the instructions, and (2) visual for the action should help reduce the split-attention effect inherent to processing complex document (Rouet, 2005). Mayer's theory of multimedia learning (2009) is based on the use of two channels. To optimize learning, the author recommends both a visual (static and/or animated images) and auditory presentation of information rather than printed text alone. This modality effect (Mousavi, Low and Sweller, 1995) has been the focus of much research showing that an audio guide is preferable to a visual guide when the environment where the tasks takes place already is a source of visual information (see Ginns, 2005). Knowledge acquisition through a virtual environment is more often assessed after training, without checking if knowledge is maintained over time (Chittaro & Buttussi, 2015). As in the first study by Ganier et al., (2013), this experiment comprises a first stage for learning the procedure and a second stage, after a few days, for recalling the procedure. The hypothesis being that the effectiveness of the virtual environment should become apparent with better performance of the learners illustrated by a learning curve evolving with the number of attempts.

2.1. Method

2.1.1. Population

Twelve student volunteers (11 men and 1 woman), with an average age of 20 (range: 19-23), participated in this study. They were all 4th year students in Brest National Engineering School (École Nationale d'Ingénieurs de Brest) and had no previous experience of the STA-R Virtuel (the virtual environment used on this study).

2.1.2. Material

We used the virtual environment for training, *STA-R Virtuel* (Le Corre, Fauvel, Hoareau, Querrec, & Buche, 2012) which is a virtual simulation of a medical laboratory (Figure 1).

INSERT FIGURE 1: Virtual Environment for training STA-R Virtuel (on the left: laboratory; on the right: touch-screen interface)

This model includes two main scenes: a lab bench on which all the chemical reagents and lab equipment (reagent bottles, pipettes, etc.) are laid out as well as a STA-R® coagulometer. A coagulometer is a device used for a number of diagnoses in haematology. It performs chronometric and immunologic tests in a limited time. The coagulometer is connected to a computer giving access to the device via a touch-screen tablet (Figure 2).

INSERT Figure 1 : STA-R® coagulometer

For this experiment, the virtual environment was presented on a laptop computer. The participant could move around from a fixed viewpoint to another in the virtual environment by using shortcuts. He/she could also interact with objects using the mouse. Audio guides could be accessed as necessary via a media player icon at the bottom left corner of the screen. The instructions for each action to be completed could be played as many times as needed. A touch-screen tablet reproducing the software used to operate the coagulometer complemented this virtual environment. A network connected both programs (i.e. the virtual environment and the software).

2.1.3. Research protocol

There were two sessions during this study. During the first session, participants were given a demonstration on how to manipulate the interface based on a different procedure than to acquire, created solely for that purpose. This procedure included all the basic interactions possible with the virtual environment. They could then ask the experimenter with any questions about the demonstration. Once the demonstration was over, participants could then take place in front of their work station made up of the laptop computer and the touch-screen tablet. Four participants performed the experiment simultaneously in the same room but communication was not allowed. The experiment consisted in performing seven consecutive attempts of the procedure for running blood tests composed of 125 actions. The procedure had been chosen in consultation with the team of specialized instructors from STAGO company and had been tested during a preliminary study. The number of attempts had been chosen on the basis of results from Ganier et al. (2013) which had showed that a simple and shorter procedure could be learnt in five or six attempts. This first session lasted about 90 minutes. After a week, participants had to take part in the second session of the experiment, during which they had to carry out three consecutive attempts of the same procedure. The same material and procedure were used in the second session. The same behavioural measures were collected. The aim was to ensure that the procedure completed the week before was memorized. This could testify to the internal validity of the virtual environment.

2.1.4. Collected data

Behavioural measures were used to assess procedure learning. For each attempt, objective performance data included the total time spent to complete the procedure, the number of referrals to the instructions and the number of incorrect actions. These measures were recorded in real time by the virtual environment under “log” files.

For ease of result processing, temporary data were converted in minutes and hundredths of minutes.

2.2. Results

Due to a transcription problem, the files corresponding to the number of incorrect actions could not be used. We will present here the results for the total time spent to complete the procedure and the number of referrals to the instructions.

2.2.1. Total time spent to complete the procedure

The time spent by participants to complete the procedure in full (Figure 3) significantly decreased with the number of attempts: F (9.99) = 98.40; p <.001. Regarding the first session (attempts 1 through 7), the time spent to complete the action decreased all along the seven attempts. Analytical pairwise comparisons show significant differences from attempt 1 to attempt 5, then a non-significant difference for the remaining attempts. This difference is particularly high between the first (22.48 min) and the second (10.60 min) attempt: F (1; 11) = 10.96; p <.001. It is lower between the next attempts (attempt 2 vs attempt 3: F (1; 11) = 7.16; p <.001; attempt 3 vs attempt 4: F (1; 11) = 2.82; p <.03; attempt 4 vs attempt 5: F (1; 11) = 5.47; p <.001). These results indicate an improvement in performance up until the 5th attempt (7 min), than a stable performance in attempts 6 (6.81 min) and 7 (6.55 min). Regarding the second session (attempts 8 through 10), analytical pairwise comparisons show significant differences between attempts 8 and 9: F (1; 11) = 3.85; p <.01, and attempts 9 and 10: F (1; 11) = 6.25; p <.001. If the learners' performance from the first and second session are compared, analytical pairwise comparisons show a significant difference between the time spent to complete the procedure between attempts 7 and 8: F (1; 11) = 4, 46; p<.001; but none during the last attempt of the first session (attempt 7) and the time spent during the last attempt of the second session (attempt 10): F (1; 11) = 2.44; p <.05.

INSERT FIGURE 3: Total time spent to complete the procedure (min)

2.2.2. Number of referrals to the audio instructions

The number of referrals to the audio instructions (Figure 4) significantly decreased with the number of attempts: F (9.99) = 121.53; p <.001. Analytical pairwise comparisons show a pattern of identical results to the time spent to complete the procedure. Indeed, regarding the first session (attempt 1 through 7), the number of referrals to the instructions decreased all along the seven attempts, from 87 referrals on average for the 1st attempt to one on the 7th. Analytical pairwise comparisons show significant differences from attempt 1 to 5 (attempt 1 vs attempt 2: F (1; 11) = 8.94; p <.001; attempt 2 vs attempt 3: F (1; 11) = 6.93; p <.001; attempt 3 vs attempt 4: F (1; 11) = 5.11; p <.001; attempt 4 vs attempt 5: F (1; 11) = 3.01; p <.03).

Regarding the second session (attempts 8 through 10), analytical pairwise comparisons show a significant difference between attempt 8 and 9: F (1; 11) = 5.22; p <.001.

Participants referred to some instructions during attempt 8. If the number of referrals to the instructions at the end of the first session is compared with the beginning of the second, analytical pairwise comparisons show a significant difference between attempt 7 and 8: F (1; 11) = 4.46; p <.001; but no difference between attempt 7 and 10: F < 1. After a week, participants were able to retrieve the procedure from their long-term memory, even if for some individuals it was necessary to refer to some instructions.

INSERT FIGURE 4 : Number of referrals to the instructions

2.3. Discussion

The objective of this experiment was to assess the internal validity of a virtual environment for learning a procedure. The results of this experiment indicate that learning through this virtual environment is feasible. This is illustrated by the learning curves based on the collected objective performance data. The total time spent to complete the procedure and the number of referrals to the instructions decrease with the number of attempts. During the first session, we can observe some stabilization of the time spent by participants to complete the task during the last attempts (attempts 6 and 7). The participants' performance during the second session show high economy regarding relearning in terms of reduction of task time. After training in a virtual environment, when participants carry out the procedure, they take 2.5 less time than when they completed the procedure the first time. It would be interesting to bring to light the underlying factors of differences in performance with the number of attempts, but also to identify the factors responsible for the cognitive overload during the first attempt. Nevertheless, it would seem that for this procedure, seven attempts were not sufficient for knowledge of the procedure to be sustainably stored in the long-term memory. The need to refer to the instructions after the first attempt of the second session (attempt 8, after a week) observed in some participants could be the consequence of an insufficient number of attempts. The assessment of the internal validity of this virtual environment shows that it is possible to learn a long procedure of 125 actions, with auditory instructions, through the help of this virtual training environment. A higher number of attempts would be preferable in order to strengthen learning. This note was taken into consideration in building the research protocol for the second experiment of this study. Its aim was to complete the assessment of the virtual environment by studying its external validity. Assessing the external validity of the virtual environment is done by verifying if the acquired skills in the virtual environment can be transferred to a real situation.

3. Experiment 2: Transfer of a procedure learnt in a virtual environment to a real situation

Beyond the possibility to teach a complex procedure, the external validity of a virtual environment can be tested through the question of transferring knowledge to a real situation (Anastassova et al., 2007). The objective of this study was to assess the external validity of the virtual environment for training to run blood tests, by evaluating the participants' performance while reproducing the procedure on actual material. To this aim, the chosen research protocol followed a paradigm for comparing two experimental situations (traditional training and virtual environment training) to a control situation.

3.1. Method

3.1.1. Population

This study was conducted in a French school (Lycée Paul Eluard in Paris). Fifty-four students (45 women and 9 men) with an average age of 20 (range: 17-22) participated in the study. They were in their 1st and 2nd year of training for Medical Biology Analyses (« BTS Analyses en Biologies Médicales »). They were divided into three groups, equivalent in number and education: traditional training, virtual environment training and control group.

3.1.2. Material

The same virtual environment (*STA-R Virtuel*) was used as in the first experiment.

It was made up of a laptop computer with a model of a medical laboratory (lab bench + coagulometer) and a touch-screen tablet with a model of the software used to operate the coagulometer.

In order to compare virtual training to a real situation, this study used actual apparatus equivalent to the models: a laboratory equipped with two lab benches on which the lab equipment (reagent bottles, pipettes, etc.) was laid out as well as a two coagulometers with a touch-screen interface (Figure 5)

Actual material was also used for reproducing the procedure further on in the training.

INSERT FIGURE 5: Schematization of the real laboratory with lab bench and coagulometer

The lab was partitioned into two separate spaces with a screen. Each space included a lab bench, a coagulometer and an area dedicated to consulting the technical document with all the actions to perform to successfully complete the procedure (Figure 6). No communication was allowed between participants who could not see each other because of the separation screen.

INSERT FIGURE 6: Transfer of the procedure on a real environment by two students

3.1.3. Research procedure

The aim for all participants was to run blood tests in the real laboratory. This procedure was the same as the one with 125 actions used in the first experiment. There were two test groups ("traditional training" and "virtual environment training") and a control group (Figure 7).

INSERT FIGURE 7: Experimental protocol of the experiment

For participants in the "traditional training" group, the training day was divided as follows: training in the morning and transfer of the procedure in the afternoon. Theoretical training covered haemostasis and the main steps of the procedure for running blood tests. This training, a projected slideshow with commentaries, was given by an instructor from STAGO company in a classroom. Later, in the workshop, the instructor presented the material to be manipulated ("This is a pipette", etc.). At this stage of training, participants had not yet manipulated the actual material. The morning training session ended with a practical training during which the instructor performed the procedure in full, using the material on the lab bench and the coagulometer. Participants then went through the procedure in groups of three, under the supervision of an instructor (ready to intervene in case of important mistake).

In the afternoon, participants had to perform the procedure in full, individually, under supervision of a STAGO technician, in the real laboratory. They could use the technical document at their disposal at any time if they needed help to perform the procedure.

The document was laid at the end of the lab bench. Participants were told to stop all manipulation when consulting the document.

For participants in the “virtual environment training” group, the morning session was identical to the one for the “traditional training” group regarding theoretical training in the classroom and presentation of the material. Training in the virtual environment replaced practical training on the actual material, during which participants performed the procedure at least 8 times on the laptop computer and connected touch-screen tablet. This minimum number of attempts was determined by the results obtained during the first experiment. Participants had a personal workstation and went through the training in groups of nine in a classroom. No communication was allowed. Transfer on the actual apparatus in the afternoon took place in the same conditions as for participants in the “traditional training” group.

Participants in the control group had the same theoretical training and presentation of the material as the other two groups. Their morning training session did not include practical training. They performed the procedure on the actual apparatus in the same conditions as the two other test groups.

For each group, the theoretical training, the presentation of the material and the practical training times were strictly identical.

3.1.4. Collected data

Data were collected with a video recording of the completion of the procedure on the actual apparatus by the participants. Objective performance data included the successful completion of the task in the time allotted (45 minutes), the total time spent to complete the procedure, the time and number of referrals to the technical document, the number of times the technician provided help to the participants (upon request or spontaneously). The technician was free to interrupt the participant at any time to inform him/her of an incorrect action if it could lead to damaging the coagulometer. The technician worked blind, as he/she was not aware which group the participant belonged to. Time counts were collected from the video recordings. For ease of statistical analysis, the chronometric data were converted in minutes and hundredths of minutes.

3.2. Results

The results regarding the successful completion of the procedure in the allotted time, then those regarding the objective performance data in the three groups are presented here.

3.2.1. Successful completion of the procedure

A time limit of 45 minutes was set to complete the procedure on the actual apparatus. This time had been determined by multiplying by three the time needed by an expert to complete the procedure at normal speed. 100% of the participants of both the “traditional training” and “virtual environment training” groups were able to complete the procedure in the allotted time. 72% of the participants in the control group successfully completed the procedure in 45 minutes. The remaining analyses were made on all participants, regardless of success or failure to complete the task in the allotted time. This data only relates to the actions taking place during the default time of 45 minutes. Beyond this time limit, the participants' actions were not taken into account.

3.2.2. Total time spent to complete the procedure and time spent consulting the technical document

Time spent to complete the procedure varied according to the training conditions (Figure 8): F (2.51) = 46.60; p <.001. The analytical pairwise comparisons indicate that participants trained in the virtual environment (m = 30.03 min; σ = 8.18) took less time to complete the procedure than participants in the control group (m = 39.09 min; σ = 5.44): F (1; 34) = 3.98; p <.05. But they took more time than the participants trained traditionally (m = 19.48 min; σ = 2.35): F (1; 34) = 24.71; p <.001. The participants with traditional training also took less time than participants in the control group: F = (1; 34) = 8.61; p <.01.

Time spent consulting the technical document during the completion of the procedure varied according to the training conditions: F (2.51) = 210.92; p <.001. The analytical pairwise comparisons show that the difference between "traditional training" (m = 0.15 min; σ = 0.32) and "virtual environment training" (m = 0.42 min; σ = 0.82) groups is not significant: F (1; 34) = 3.17; ns. Participants in the control group (m = 8.55 min; σ = 2.25) spent a longer time referring to the technical document than participants with "traditional training" F (1; 34) = 16.13; p <.001; and participants with "virtual environment training": F (1; 34) = 8.99; p <.01.

INSERT FIGURE 8 : Total time spent to complete the procedure (min) with time of completion and time spent reading the technical document

3.2.3. Number of referrals to the technical document

The number of referrals to the technical document varied according to the experimental conditions (Figure 9): F (2.51) = 232.93; p <.001. As for the time spent consulting the technical document, analytical pairwise comparisons indicate that the number of referrals to the technical document is significantly less elevated for the “traditional training” (m = 0.7; σ = 1.49): F (1; 34) = 21.43; p <.001; and the “virtual environment training” (m = 1.89; σ = 4.68): F (1; 34) = 11.98; p <.001; groups than for the control group (m = 48.72 ; σ = 12.24). There is no significant difference between both groups with training: F (1; 34) = 1.95; ns.

INSERT FIGURE 9: Number of referrals to the technical document

3.2.4. Number of times the technician provided help upon request

The number of times the technician provided help upon request varied according to the training conditions: F (2.51) = 36.48; p <.001. Analytical pairwise comparisons show that participants trained in the virtual environment (m = 0.7; σ = 0.7) requested the help of the technician more frequently than those with traditional training (m = 0.2; σ = 0.4): F (1; 34) = 7.07; p <.03. But less so than the participants from the control group (m = 3.1; σ = 1.7): F (1; 34) = 10.45; p <.01.

3.2.5. Number of times the technician provided help spontaneously

The number of times the technician provided help spontaneously varied according to the training conditions: F (2; 51) = 5.47; p <.01. The technician intervened more frequently with the control group (m = 3.9; σ = 2.8) than with the "traditional training" (m = 1.4; σ = 1.8) and "virtual environment training" (m = 2.3; σ = 2.2) groups with no significant difference. Similarly, there is no significant difference between the two groups with training. If both groups with training are considered together, analytical pairwise comparisons indicate that the technician intervened more frequently with the control group than with the other two groups: F (1; 51) = 1.26; p <.01.

3.3. Discussion

This experiment was designed to assess the external validity of a virtual environment for learning a procedure. The results indicate that a transfer does occur; as for the participants who were trained traditionally, all the participants who were trained with the virtual environment successfully completed the procedure in the allotted time. However, transfer does not take place in the same conditions according to the type of training. If both groups with training are compared, the only significant difference is found in the total time taken to complete the procedure. Indeed, the participants in the "virtual environment training" group took 35% additional time to complete the procedure than those in the "traditional training" group. This difference could be explained by the design choice of the virtual environment (for example, a lack of similarity between the virtual and real software used to operate the machine) but also by the difficulty to appreciate the actual time needed for the coagulometer to perform the tests. In real conditions, after the blood tests begin, the coagulometer comes to life (pistons movement, noise, etc.) for a period of time depending on the tests required and when it stops, the rest of the procedure can then be completed.

Participants trained with the virtual environment had a difficult time evaluating when the coagulometer was done with the tests. These design defaults can be identified by conducting an ergonomic analysis of the interface.

4. General discussion

The objective of this study was to test a virtual environment built for learning a procedure for running blood tests on a coagulometer. This was achieved through two experiments. The first one was designed to assess the internal validity of the virtual environment by checking the effectiveness of procedural learning through training in the virtual environment. In order to assess the external validity of the virtual environment, the second experiment addressed the question of transfer of the acquired skills from a virtual environment to real conditions. The results of the first experiment show that it is possible to learn a procedure with a virtual environment. However, after a week, participants must refer to some instructions again in order to complete the procedure in full. In addition, their performance further increases during this second session. This means that the virtual environment as designed makes learning possible but not optimal. Factors to facilitate learning from the first attempt must now be identified. Making the first attempt easier would have an impact on learning as a whole.

The results of the second experiment indicate that the participants trained in the virtual environment perform the procedure better than the participants in the control group. However, even if they did not succeed in training as well as the participants with traditional training, their performance was not completely unsuccessful. This could be explained by the fact that for participant trained in the virtual environment didn't manipulates the real object. The use of mixed reality (VR and AR) can then be envisage. This difference could be also the consequence of an insufficient or not fully conclusive learning in the virtual environment. Facilitating the information processes involved in training in the virtual environment, and more precisely the first stage of learning (here the first attempt), should have an impact on the ability to transfer learning to a real situation.

The virtual environment used in both experiments is similar to a large number of virtual environments used in training in terms of interaction (keyboard & mouse) or design of the training scenario. It allows to performe a set of actions (short or long) on the same hierarchy level, i.e., undistinguishing titles, subtitles and low level instructions by typography or layout marks (Bliss et al., 1997; Ganier et al., 2014; Hapeshi & Jones, 1992; Legendre, Sahmoune Rachedi, Descamps, & Fernandez, 2014; Moskaliuk et al., 2013). This study aimed to assess the internal and external validity of a virtual environment for learning a procedure. The results provide trends for both. However, learning and transfer performance can still be improved. The objective now lies in bringing modifications to the existing environment and to test their impact on procedure learning, and transfer to real conditions. Envisioned improvements may take several directions, for example, regarding cognitive processes employed when learning. Participants experience the most difficulties during the first attempt. The cognitive cost of this first attempt could be reduced by adding one or several forms of guidance for the learner, or by modifying the content in the instructions. The exploration of the cognitive process could be enriched by collecting self-reported data to evaluate the difficulties encountered by the learners. Improvements could also focus on the user by questioning the forms of presence (immersion/interaction) with the environment to compare immersive display (i.e. headset or CAVE) versus non-immersive display (i.e. desktop VR) for example or by taking into account interindividual differences. The sequence of learning could be personalized according to the learners' abilities and their progression during training, for example by selecting information and help available all along the training. Implementing these improvements could be assessed by reproducing the two experiments presented here in order to compare the "improved" virtual environment with the first draft used in this study, in a test-retest paradigm.

Conflict of interest: none.

References

Anastassova, M., Burkhardt, J.-M., Mégard, C., & Ehanno, P. (2007). L'ergonomie de la réalité augmentée pour l'apprentissage: une revue. *Le Travail Humain*, *70*(2), 97–125.

Anderson, J. (1982). Acquisition of Cognitive Skill. *Psychological Review*, *89*(4), 369–406.

Anderson, J. (1998). Instructional Applications of ACT : Past , Present , and Future. In *ITS'98* (p. 1). San Antonio, TEX.

Anderson, J. (2013). *The architecture of cognition* (2nd ed.). New York: Psychology Press.

Anderson, J., Reder, L., & Simon, H. (1996). Situated learning and education. *Educational Researcher*, *25*(4), 5–11.

Annetta, L., Minogue, J., Holmes, S., & Cheng, M. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education*, *53*(1), 74–85.

Arnaldi, B., Fuchs, P., & Tisseau, J. (2003). Introduction. In *Traité de la réalité virtuelle. Volume 1 : L'homme et l'environnement virtuel* (pp. 3–21). Les Presses de l'Ecole de Mines de Paris.

Beyer-Berjot, L., Palter, V., Grantcharov, T., & Aggarwal, R. (2014). Advanced training in laparoscopic abdominal surgery: A systematic review. *Surgery*, *156*(3), 676–688.

Bliss, J., Tidwell, P., & Guest, M. (1997). The effectiveness of virtual reality for administering spatial navigation training to firefighters. *Presence*, *6*(1), 73–86.

Bossard, C., Kermarrec, G., Buche, C., & Tisseau, J. (2004). Transfer of learning in Virtual Environments : a new challenge ? *Virtual Reality*, *12*(3), 151–161.

Boud, A. C., Baber, C., & Steiner, S. J. (2000). Virtual Reality : a tool for assembly ? *Presence*, *9*(5), 486–496.

Bracke, D. (1998). Vers un modèle théorique du transfert: les contraintes à respecter. *Revue Des Sciences de L'éducation*, *24*(2), 235–266.

Bransford, J. D., & Schwartz, D. L. (2001). Rethinking Transfer : A Simple Proposal With Multiple Implications. *Research in Education*, *24*(3), 61–100.

Brooks, F. P. (1999). What's Real About Virtual Reality ? *IEEE Computer Graphics and Application*, *19(6)*(December), 16–27.

Bruwaene, S. Van, Schijven, M., & Miserez, M. (2014). Assessment of Procedural Skills Using Virtual Simulation Remains a Challenge. *Journal of Surgical Education*, *71*(5), 654–661.

Butterfield, E., & Nelson, G. (1989). Theory and practice of teaching for transfer. *Educational Technology Research and Development*, *37*(3), 5–38.

Carlson, P., Peters, A., Gilbert, S., Vance, J. M., & Luse, A. (2015). Virtual Training: Learning Transfer of Assembly Tasks. *IEEE Transactions on Visualization and Computer Graphics*, *2626*(c), 1–14.

Carpentier, K., & Lourdeaux, D. (2013). Sélection dynamique de situations d'apprentissage en environnement virtuel. In *Environnements Informatiques pour l'Apprentissage Humain* (pp. 129–140). Toulouse, France.

Cazeaux, E., Devillers, F., & Saint-Romas, C. (2005). *Giat virtual training formation a la maintenance*. *Laval Virtual*. Laval.

Chen, J. A., Tutwiler, M. S., Metcalf, S. J., Kamarainen, A., Grotzer, T., & Dede, C. (2016). A multi-user virtual environment to support students' self-efficacy and interest in science: A latent growth model analysis. *Learning and Instruction*, *41*, 11–22.

Chittaro, L., & Buttussi, F. (2015). Assessing Knowledge Retention of an Immersive Serious Game vs . a Traditional Education Method in Aviation Safety. *IEEE Transactions on Visualization and Computer Graphics*, *21*(4), 529–538.

Dalgarno, B., & Lee, M. (2010). What are the learning affordances of 3- D virtual environments? *British Journal of Educational Technology*, *41*(1), 10–32.

Dede, C., Salzman, M., Loftin, R., & Ash, K. (1997). Using virtual reality technology to convey abstract scientific concepts. In M. Jacobson & R. Kozma (Eds.), *Learning the sciences of the 21st Century: Research, Design, and Implementing Advanced Technology Learning Environments* (pp. 361–413). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dixon, P., Harrison, K., & Taylor, D. (1993). Effects of sentence form on the construction of mental plans from procedural discourse. *Canadian Journal of Experimental Psychology*, *47*(2), 375.

Dorn, L., & Barker, D. (2005). The effects of driver training on simulated driving performance. *Accident; Analysis and Prevention*, *37*(1), 63–9.

El-Chaar, J., Boer, C. R., Pedrazzoli, P., Mazzola, S., & Maso, G. D. (2011). Interactive 3D virtual environments for industrial operation training and maintenance. In *ICRMS'2011 - Safety First, Reliability Primary: Proceedings of 2011 9th International Conference on Reliability, Maintainability and Safety* (pp. 1376–1381). Guiyang.

Fuchs, P. (1996). *Les interfaces de la réalité virtuelle*. Paris: Les Presses de l'Ecole de Mines de Paris.

Gallagher, A., & Cates, C. (2004). Approval of virtual reality training for carotid stenting: what this means for procedural-based medicine. *Jama*, *292*(24), 3024–3026.

Gallagher, A., Seymour, N., Jordan-Black, J., McGlade, K., & Satava, R. (2013). Prospective, randomized assessment of transfer of training (ToT) and transfer effectiveness ratio (TER) of virtual reality simulation training for laparoscopic skill. *Annals of Surgery*, *257*(6), 1025–1031.

Ganier, F. (2004). Factors affecting the processing of procedural instructions: implications for document design. *IEEE Transactions on Professional Communication*, *47*(1), 15–26.

Ganier, F., Hoareau, C., & Devillers, F. (2013). Évaluation des performances et de la Charge de travail induits par l'apprentissage de procédures de maintenance en Environnement Virtuel. *Le Travail Humain*, *76*(4), 335–363.

Ganier, F., Hoareau, C., & Tisseau, J. (2014). Evaluation of procedural learning transfer from a virtual environment to a real situation : a case study on tank maintenance training. *Ergonomics*, *57*(6), 828–843.

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: How effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, *29*(3), 207–219.

Greunke, L., & Sadagic, A. (2016). Taking Immersive Leap in Training of Landing Signal Officers. *IEEE Transactions on Visualization and Computer Graphics*, *22*(4), 1482–1491.

Grumbach, A., & Klinger, E. (2007). Virtuel et Cognition. *Intellectica*, *1*(45), 7–22.

Hale, K., & Stanney, K. (2015). *Handbook of virtual environments: Design, implementation, and applications*. (K. Hale & K. Stanney, Eds.) (CRC Press). Boca Raton, FL.

Hapeshi, K., & Jones, D. (1992). Interactive multimedia for instruction: A cognitive analysis of the role of audition and vision. *International Journal of Human-Computer Interaction*, *4*(1), 79–99.

Haskell, R. (2000). *Transfer of learning: Cognition, Instruction, and Reasoning*. London: Academic Press.

Hasselbring, T., & Sherwood, R. (1987). An evaluation of a level-one instructional videodisc program. *Journal of Educational Technology Systems*, *16*(2), 151–169.

Heurley, L. (1997). Vers une définition du concept de texte procédural: le point de vue de la psycholinguistique. *Les Cahiers Du Français Contemporain*, *4*, 109–133.

Heurley, L., & Ganier, F. (2006). L â€™ utilisation des textes procéduraux : lecture , compréhension et exécution d â€™ instructions écrites. *Intellectica*, *2*(44), 1–18.

Hoareau, C., Ganier, F., Querrec, R., Corre, F. Le, Buche, C., Le Corre, F., & Buche, C. (2013). Evolution of cognitive load when learning a procedure in a Virtual Environment for Training. In *International Cognitive Load Theory Conference* (pp. 130–132). Toulouse.

Huegel, J. C., Celik, O., Israr, A., & O'Malley, M. K. (2009). Expertise-based performance measures in a virtual training environment. *Presence: Teleoperators and Virtual Environments*, *18*(6), 449–467.

Jorge, V. A. M., Sarmiento, W. J., Maciel, A., Nedel, L., Collazos, C. A., Faria, F., & Oliveira, J. (2013). Interacting with danger in an immersive environment: issues on cognitive load and risk perception. In *19th ACM Symposium on Virtual Reality Software and Technology - VRST '13* (pp. 83–92). Singapore.

Jou, M., & Wang, J. (2013). Investigation of effects of virtual reality environments on learning performance of technical skills. *Computers in Human Behavior*, *29*(2), 433–438.

Kho, R. (2011). Comparison of robotic-assisted laparoscopy versus conventional laparoscopy on skill acquisition and performance. *Clinical Obstetrics and Gynecology*, *54*(3), 376–381.

Kieras, D. E. DE, & Bovair, S. (1986). the acquisition of procedures from text : a production-system analysis of transfer of training. *Journal of Memory and Language*, *25*(5), 507–524.

Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Ergonomics Science*, *14*(1), 22–37.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: effect of direct instruction and discovery learning. *Psychological Science*, *15*(10), 661–7.

Kozak, J., Hancock, P., Arthur, E., & Chrysler, S. (1993). Transfer of training from virtual reality. *Ergonomics*, *36*(7), 777–784.

Lam, C. K., Sundaraj, K., & Sulaiman, M. N. N. (2014). A Review of Computer-Generated Simulation in the Pedagogy of Cataract Surgery Training and Assessment. *International Journal of Human-Computer Interaction*, *29*(10), 661–669.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge, UK: Cambridge University Press.

Le Corre, F., Fauvel, C., Hoareau, C., Querrec, R., & Buche, C. (2012). Chrysaor: an agent-based intelligent tutoring system in virtual environment. In *International Conference on Virtual Learning* (pp. 39–45). Brasov.

Lee, A. (1998). Transfer as a measure of intellectual functioning. *Perspectives on Fundamental Processes in Intellectual Functioning: A Survey of Research Approaches*, *1*, 351–366.

Lee, A., & Pennington, N. (1993). Learning computer programming: A route to general reasoning skills. In C. R. Cook, J. C. Scholtz, & J. C. Spohrer (Eds.), *Empirical studies of programmers: Fifth wokshop* (pp. 113–136). Norwood, NJ: Ablex Publishing Corporation.

Legendre, G., Sahmoune Rachedi, L., Descamps, P., & Fernandez, H. (2014). Apport d’un simulateur virtuel d’anatomie périnéale (Pelvic Mentor) dans l’apprentissage de la pelvi-périnéologie : résultats d’une étude préliminaire. *Journal de Gynecologie, Obstetrique et Biologie de La Reproduction*, *44*(1), 72–

77.

Littlefield, J., Delclos, V., Bransford, J., Clayton, K., & Franks, J. (1989). Some prerequisites for teaching thinking: Methodological issues in the study of LOGO programming. *Cognition and Instruction*, *6*(4), 331–366.

Loukas, C., Nikiteas, N., Kanakis, M., & Georgiou, E. (2011). Evaluating the effectiveness of virtual reality simulation training in intravenous cannulation. *Simulation in Healthcare : Journal of the Society for Simulation in Healthcare*, *6*(4), 213–217.

Loup-Escande, É., Burkhardt, J.-M., & Richir, S. (2013). Anticiper et évaluer l'utilité dans la conception ergonomique des technologies émergentes : une revue. *Le Travail Humain*, *76*(1), 27–55.

Masciocchi, C., Dark, V., & Parkhurst, D. (2007). *Evaluation of virtual reality snowplow simulation training*. Iowa State University.

McKeown, M., & Beck, I. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Literacy Research*, *15*(1), 3–18.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers and Education*, *70*, 29–40.

Miles, H., Musembi, N., Pop, S. R., & John, N. W. (2013). Investigation of a Virtual Environment for Rugby Skills Training. In *International Conference on Cyberworlds* (pp. 56–63). Yokohama.

Moskaliuk, J., Bertram, J., & Cress, U. (2013). Training in virtual environments: putting theory into practice. *Ergonomics*, *56*(2), 195–204.

Oren, M., Carlson, P., Gilbert, S., & Vance, J. M. (2012). Puzzle assembly training: Real world vs. virtual environment. In *IEEE Virtual Reality* (pp. 27–30). Iowa.

Paiva, P. V. D. F., Machado, L. S., & Valenca, A. M. G. (2013). A Virtual Environment for Training and Assessment of Surgical Teams. In *2013 XV Symposium on Virtual and Augmented Reality* (pp. 17–26). Cuiabá.

Park, C.-H., Jang, G., & Chai, Y.-H. (2006). Development of a Virtual Reality Training System for Live-Line Workers. *International Journal of Human-Computer Interaction*, *20*(3), 285–303.

Pennington, N., Nicolich, R., & Rahm, J. (1995). Transfer of Training between Cognitive Subskills : Is Knowledge Use Specific ? *Cognitive Psychology*, *28*, 175–224.

Perkins, D., & Salomon, G. (1992). Transfer of Learning. In *International Encyclopedia of Education* (pp. 2–15). Oxford, England.

Presseau, A. (2000). Analyse de l ' efficacité d ' interventions sur le transfert des apprentissages en mathématiques. *Revue Des Sciences de L'éducation*, *XXVI*(3), 515–544.

Rauter, G., Sigrist, R., Koch, C., & Crivelli, F. (2013). Transfer of complex skill learning from virtual to real rowing. *PLoS ONE*, *8*(12), 1–18.

Raynal, F., & Rieunier, A. (2012). *Pédagogie: Dictionnaire des concepts clés: Apprentissages, formation, psychologie cognitive* (9ème). Issy-les-Moulineaux:

ESF éditeur.

Rose, F., Attree, E., Brooks, B. M., Parslow, D. M., Penn, P. R., & Ambihaipahan, N. (1998). Transfer of training from virtual to real environments. In *2nd European Conference on Disability, Virtual Reality and Associated Technologies* (pp. 69–75). Skövde.

Rose, F., Brooks, B. M., Parslow, D. M., Penn, P. R., Ambihaipahan, N., Attree, E., … Penn, P. R. (2000). Training in virtual environments : transfer to real world tasks and equivalence to real training. *Ergonomics*, *43*(4), 494–511.

Rouet, J.-F. (2005). La conception des ressources multimédias pour l'apprentissage : apports des recherches en psychologie du langage. *Revue Française de Pédagogie*, *152*, 79–87.

Singley, M., & Anderson, J. (1987). A keystroke analysis of learning and transfer in text editing. *Human-Computer Interaction*, *3*(3), 223–274.

Spruit, E. N., Band, G. P. H., Hamming, J. F., & Ridderinkhof, K. R. (2014). Optimal training design for procedural motor skills: a review and application to laparoscopic surgery. *Psychological Research*, *78*(6), 878–891.

Tardif, J. (1999). Le transfert des apprentissages.

Thorndike, E., & Woodworth, R. (1901). The influence of improvement in one mental function upon the efficiency of other functions: III. Functions involving attention, observation and discrimination. *Psychological Review*, *8*(6), 247–261.

Wenger, E., & Lave, J. (1991). *Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives) by*.

Cambridge, UK: Cambridge University Press.

Witmer, B., & Singer, M. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240.

Wrzesien, M., & Raya, M. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers & Education*, *55*(1), 178–187.

Author biographies

Charlotte Hoareau obtained her Ph.D. in Cognitive Psychology and Ergonomics at the University of Brest in 2016. She is a member of the IHSEV team of the Lab-STICC laboratory. She is specialized in procedural learning, virtual environment for training and assessment of new interfaces as well as virtual environments.

Ronan Querrec received a habilitation to direct researches degree in 2010 and a PhD degree in 2001. He is a member of the board of director of the French

Association for Virtual Reality (AFRV). He is at the head of the CERV (Research Center for Virtual Reality).

Cédric Buche received a habilitation to direct researches degree in 2012 and a PhD degree in 2005. He is a member of the board of director of the French Association for Artificial Intelligence (AFIA). He is the Editor-in-Chief of the International Journal of Virtual Reality (IJRV).

Franck Ganier is a professor of Cognitive Psychology and Ergonomics at the University of Brest, France. He is a member of the Lab-STICC Laboratory, European Centre for Virtual Reality. His research interests concern procedural learning, processes for comprehending instructions and methods for evaluating technical documents as well as virtual environments.

Table of Responses to IJCHI Reviewers' Comments

Reviewer #	Reviewer Issue/Comment to Be Addressed	Response/Action Taken	Document Location
1	Some more recent VR literature has been added (three papers were added), however the bulk of VR references is still dated and major VR literature is missing.	We have revised our definition of VR and a part of the literature to make it more current; for example: Beyer-Berjot & al. (2014), Chen & al. (2016), Dede & al. (1997), Hale & Stanney (2015), Lam & al. (2014), Huegel & al. (2009), Loukas & al. (2011), Merchant & al. (2014), Wrzesien & Raya (2010),…	p.1-5
1	Page 3: "For the past fifteen years, many applications have been developed but few of them proposed a validation of its real effects on learning. The effectiveness of virtual environments for training has already been studied…"	"For the past fifteen years, many applications have been developed but few of them proposed a validation of its real effects on learning. The effectiveness of virtual environments for training has already been studied (Annetta et al., 2009; Chen et al., 2016; Dede, Salzman, Loftin, & Ash, 1997; Huegel, Celik, Israr, & O'Malley, 2009; Loukas, Nikiteas, Kanakis, & Georgiou, 2011; Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014; Wrzesien & Raya, 2010)"	p.2
1	1.3. Objective and materials —> 1.3. Objective, materials, and scope	1.3. Objective, materials and scope	p.12
2	This version of manuscript has improved, but there are still several places with grammatical errors. Specific suggestions are listed as follows:	We tried to correct a maximum of grammatical errors (throughout the manuscript) in addition to the specific suggestions mentioned below.	
2	Page 4, section 1.1 first sentence. Need to provide reference for the direct quote of definition of virtual reality.	"Virtual reality can be defined (Fuchs, 1996), from a technical point of view as "a scientific and technical field using information technology and behavioural interfaces in order to simulate in a virtual world the behaviours of 3D entities, interacting in real time with each other and with one or more users in real time, through pseudo-natural immersion via sensory motor channel" (cited by Arnaldi, Fuchs, & Tisseau, 2003, p. 8). If VR was then only used in some specifics domains, nowadays with the development of technologies (coming essentially from mobile phone industries), we can envisage it use for large audiences (leisure, entertainment…). As this term is ambiguous, lots of researchers prefer to use the term of Virtual Environment (Hale & Stanney, 2015)."	p.4
2	Page 4, last sentence. "this permits to consider VR…." Awkward expression. Please revise	"This allows us to consider VR as an interesting tool for learning."	p.5
2	Page 4, last line, "it has been used in training, it can help……" grammatical error, please correct.	"It has been used in training, because-it can help simulate conditions close to reality"	p.5
2	Page 12, last sentence in second paragraph, "will be treated in future works". The word "treated" should be modified.	"Impact of level of presence (immersion/interaction) will be addressed in future works."	p.12
2	Page 12, middle of page, "the effect of split-attention inherent to processing complex document." And the reference "Mayer's theory of multimedia learning (2009) is based on the use of two channels. To optimize learning and to make up for the split-attention, the author recommends both a visual (static and/or animated images) and auditory presentation of information rather than printed text alone." Here, the word split attention is posed as negative, implying users has to divide their attention between visual and auditory presentation. However, the use of dual channel should have positive effect on	"To optimize learning, the author recommends both a visual (static and/or animated images) and auditory presentation of information rather than printed text alone. This modality effect (Mousavi, Low and Sweller, 1995) has been the focus of much research showing that an audio guide is preferable to a visual guide when the environment where the tasks takes place already is a source of visual information (see Ginns, 2005)."	p.13

	learning. Therefore, I suggest removing "split-attention" in the discussion		
2	Page 29 "It enables a set of actions to be performed, whether they be short or long, all on the same hierarchy level, i.e., undistinguishing titles, subtitles and low level instructions by typography or layout marks" The modified sentence is still confusing as to what it means.	"It allows to performed a set of actions (short or long) on the same hierarchy level"	p.29
2	The author indicated that they will collect self-reported data in future experiment. Please indicate this at the end of the article for limitation and future work.	"The objective now lies in bringing modifications to the existing environment and to test their impact on procedure learning, and transfer to real conditions. Envisioned improvements may take several directions, for example, regarding cognitive processes employed when learning. Participants experience the most difficulties during the first attempt. The cognitive cost of this first attempt could be reduced by adding one or several forms of guidance for the learner, or by modifying the content in the instructions. The exploration of the cognitive process could be enriched by collecting self-reported data to evaluate the difficulties encountered by the learners."	p.29
	Reference at least three articles already published in the International Journal of Human-Computer Interaction on this topic or on a related topic	Hapeshi, K., & Jones, D. (1992). Interactive multimedia for instruction: A cognitive analysis of the role of audition and vision. *International Journal of Human-Computer Interaction*, *4*(1), 79–99. Lam, C. K., Sundaraj, K., & Sulaiman, M. N. N. (2014). A Review of Computer-Generated Simulation in the Pedagogy of Cataract Surgery Training and Assessment. *International Journal of Human-Computer Interaction*, *29*(10), 661–669. Park, C.-H., Jang, G., & Chai, Y.-H. (2006). Development of a Virtual Reality Training System for Live-Line Workers. *International Journal of Human-Computer Interaction*, *20*(3), 285–303.	p.2 p.29

Figure1 - Virtual Environment for Training STA-R Virtuel

Figure2 - STA-R coagulometer

Click here to download Figure Figure2.tiff

Figure3 - Total time (min)

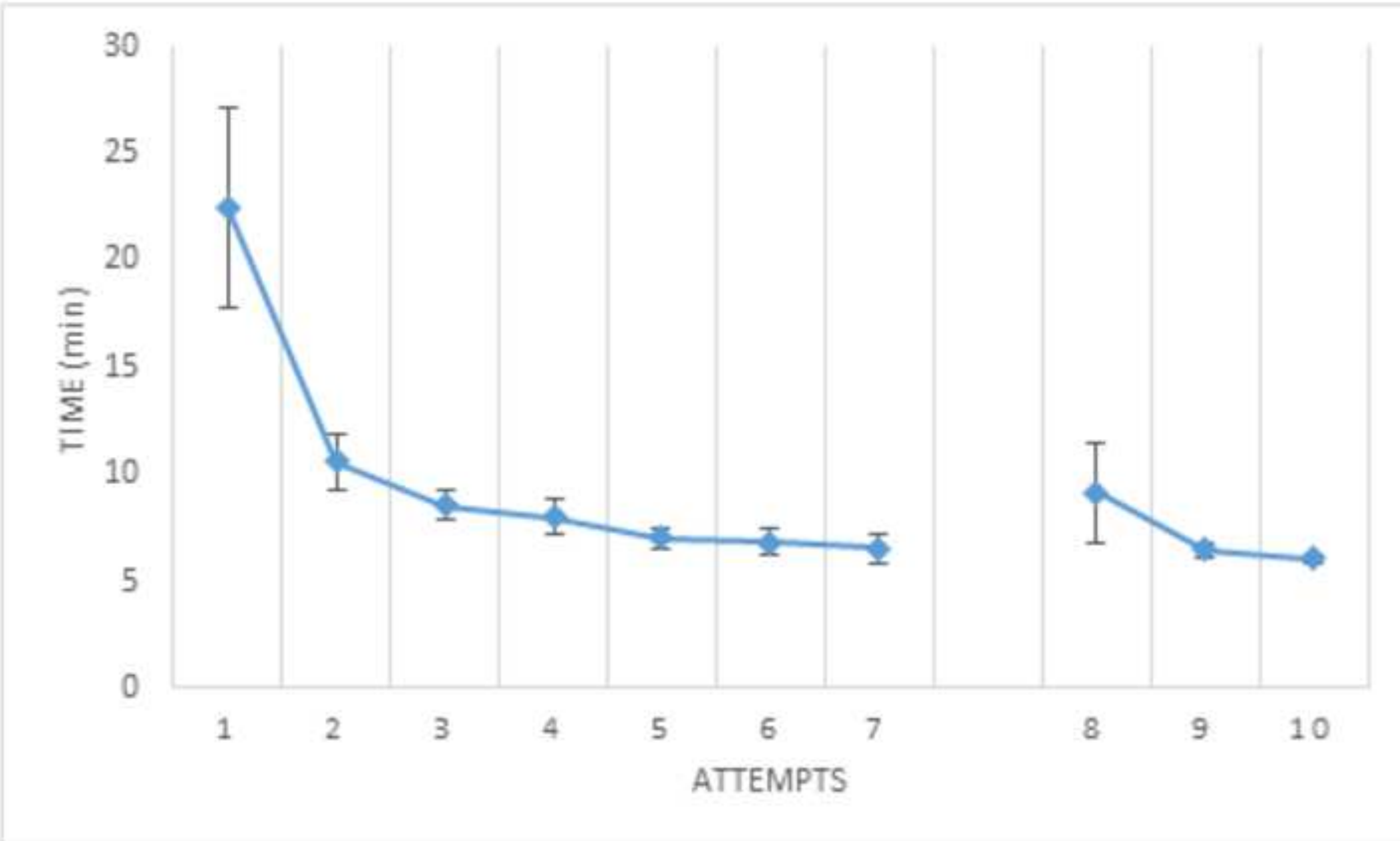

Figure4 - Number of referrals to the instructions

Click here to download Figure Figure4.tif

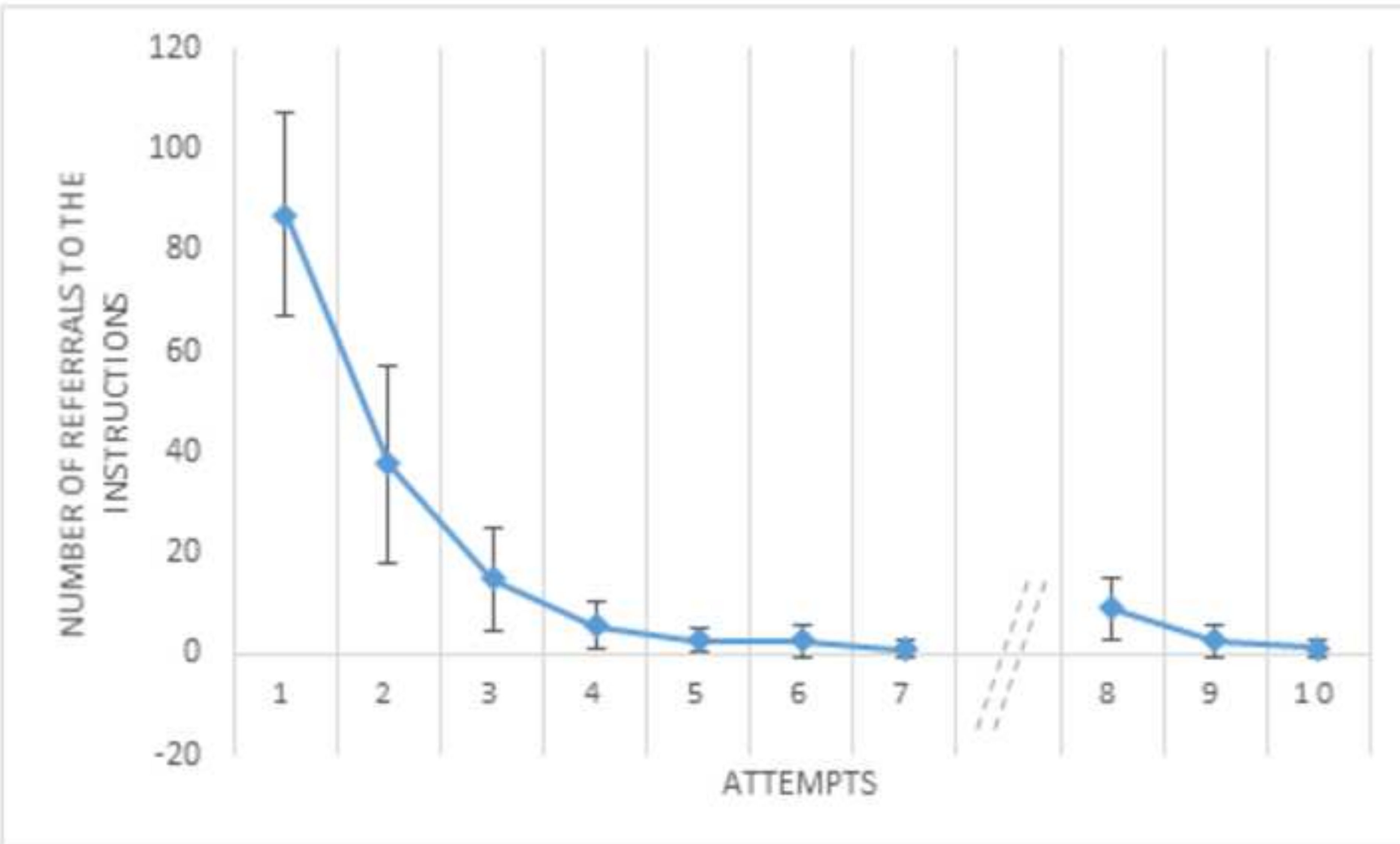

Figure5 - Schematization of the laboratory

Click here to download Figure Figure5.tif

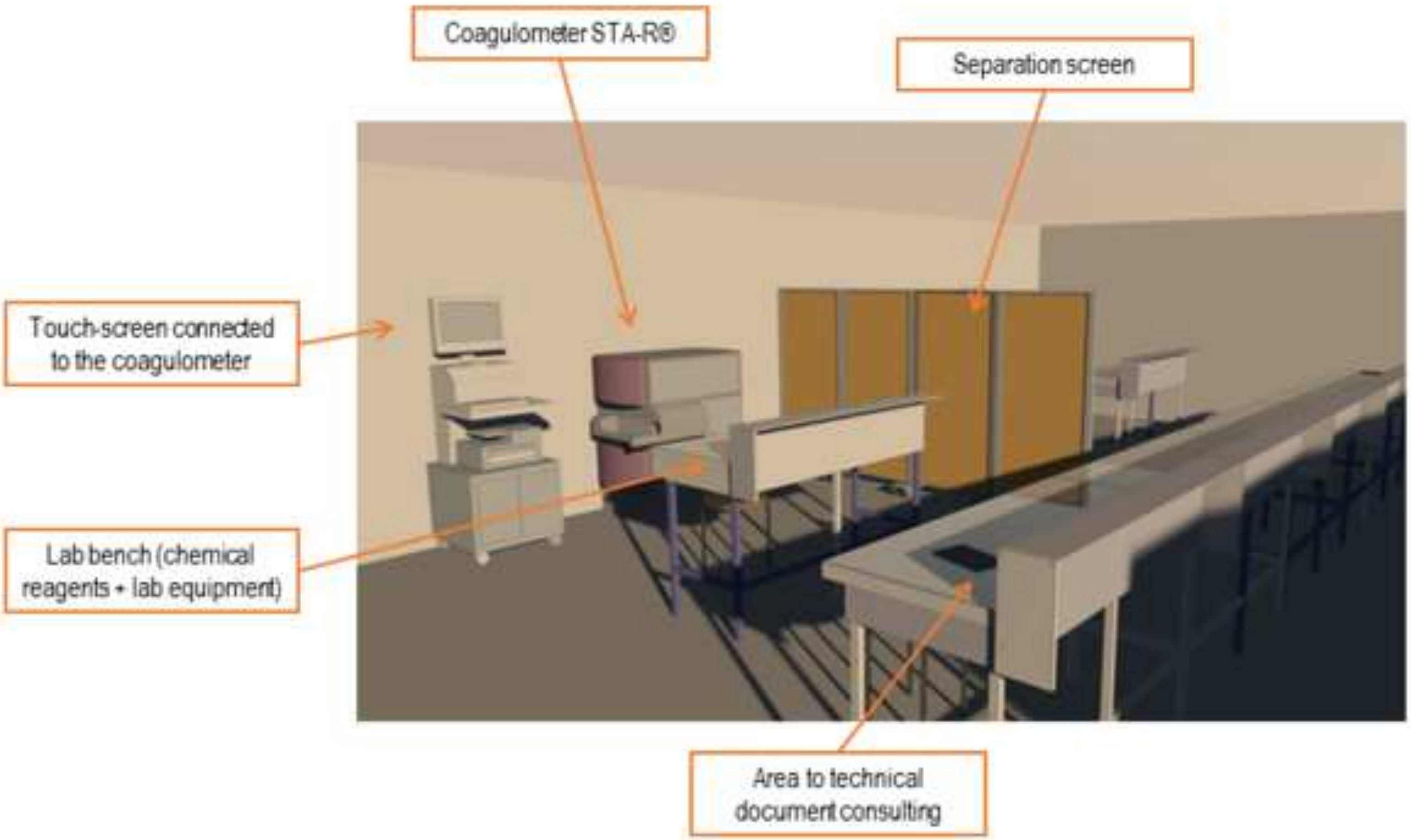

Figure6 - Transfer on a real environment

Click here to download Figure Figure6.tif

Figure7 - Experimental protocol

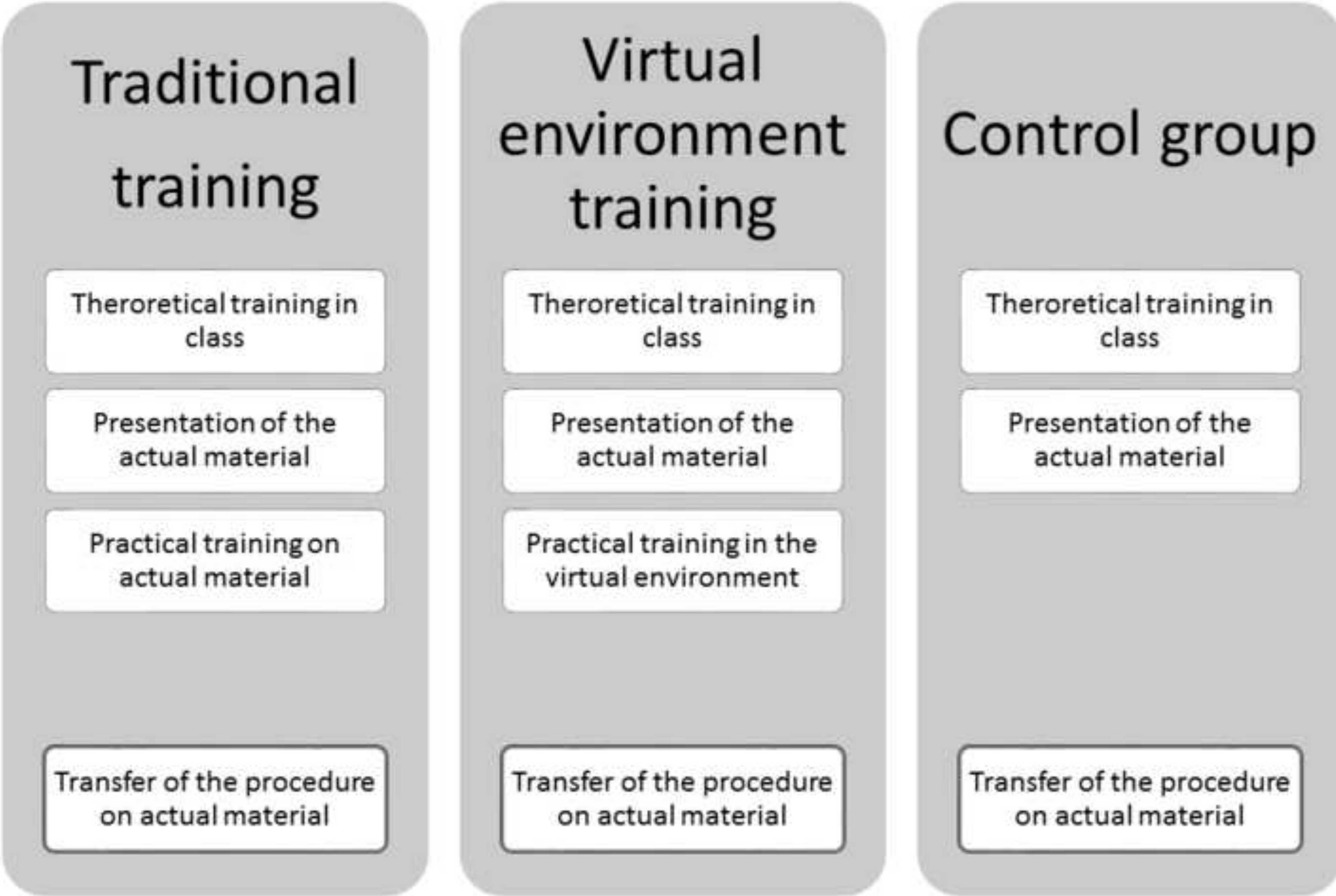

Figure8 - Total time to complete the procedure on the real device Click here to download Figure Figure8.tif

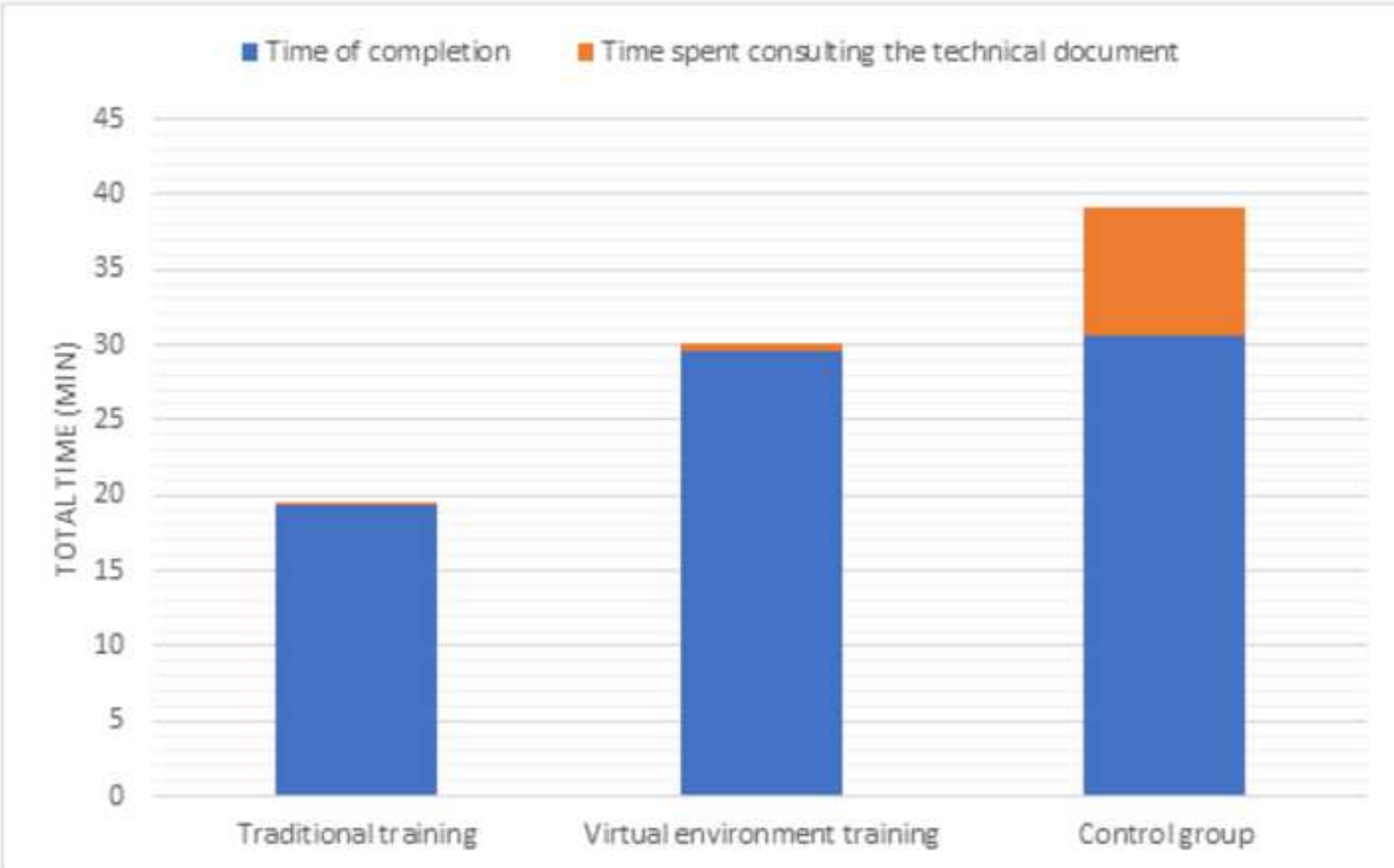

Figure9 - Number of referrals to the technical document