



HAL
open science

Ad-hoc study on soldiers calibration procedure in Virtual Reality

Jean-Daniel Taupiac, Nancy Rodriguez, Olivier Strauss, Martin Rabier

► **To cite this version:**

Jean-Daniel Taupiac, Nancy Rodriguez, Olivier Strauss, Martin Rabier. Ad-hoc study on soldiers calibration procedure in Virtual Reality. IEEE-VR 2019 - 26th Conference on Virtual Reality and 3D User Interfaces, Mar 2019, Osaka, Japan. pp.190-199, 10.1109/VR.2019.8797854 . hal-02090811

HAL Id: hal-02090811

<https://hal.science/hal-02090811v1>

Submitted on 5 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Ad-hoc study on soldiers calibration procedure in Virtual Reality

Jean-Daniel Taupiac*
LIRMM, Univ. Montpellier, CNRS,
Montpellier, France
Capgemini Technology Services,
Bayonne, France

Nancy Rodriguez†
LIRMM
Univ. Montpellier, CNRS,
Montpellier, France

Olivier Strauss‡
LIRMM
Univ. Montpellier, CNRS,
Montpellier, France

Martin Rabier§
Capgemini Technology
Services, Bayonne, France

ABSTRACT

French Army infantrymen's are equipped today with a combat system called FELIN, which includes an infrared sighting device: the IR sight. One of the first manipulations learned by the soldier is the IR sight calibration. Currently, calibration training is a two-step process. The first step consists of practicing on a 2D WIMP software until making no mistakes. Then, the soldiers can apply his knowledge in the real situation on the shooting range. In this paper, we present an ad-hoc study of a learning method including a prototype in Virtual Reality for training on the FELIN IR sight calibration procedure. It has been experimented on real infantrymen learners in an infantry school. Results showed an attractive added value of Virtual Reality in this specific use case. It improved the learners' intrinsic motivation to repeat the training task as well as the learning efficiency. It also helped the training team to identify specific mistake types not detected by the traditional learning software.

1 INTRODUCTION

French infantrymen are equipped with an infrared sight device called IR sight, within their combat system FELIN (*Fantassin à Équipement et Liaisons INTégrés*: Integrated Infantryman Equipment and Communications).

One of the first main tasks a new infantryman has to learn is the IR sight calibration. This technical procedure requires him to apply his knowledge about the sight system and calculate the appropriate correction while staying focused on his senses during the control shooting phases. Learning is supported by a 2D WIMP software called EAO FELIN, which allows a simultaneous training of a large number of learners. This software shows some learning limits. When questioned, instructors say that learners must make too many replays before making no mistakes. Moreover, they suppose that replaying the task without analyzing the mistakes might distort the procedure understanding. They also underline potential efficiency lack for skill transfer when learners are on the shooting range: learners seem to spend time on their first procedures.

As part of a collaborative work between the French Land Army, Capgemini and the LIRMM (Montpellier Laboratory of Informatics, Robotics and Microelectronics), we are interested in measuring the contribution of Virtual Reality for the IR sight calibration training procedure. Capgemini is also currently carrying out a long-term application management of the EAO FELIN 2D software.

This paper is organized as follows. In the first section, we present related work. Afterward, we present the FELIN combat system, its IR sight, the calibration procedure and the *EAO FELIN* learning software traditionally used. In section 4, we present the VR prototype we implemented, and the experiments we carried out. We end this paper by analyzing and underlying the tracks of our future works.

*e-mail: jean-daniel.taupiac@lirmm.fr

†e-mail: nancy.rodriguez@lirmm.fr

‡e-mail: olivier.strauss@lirmm.fr

§e-mail: martin.rabier@capgemini.com

2 RELATED WORK

Virtual Reality was used for technical procedures training in numerous fields: maintenance procedures [23], medical procedures [46], emergency procedures [37]. In military area, it also has been used in technical procedures [25] as well as in fight training [36], security procedures [53] and social skills [26]. Even though Virtual Reality has been used for marksmanship training [19,42], we found no works in the literature concerning rifle maintenance procedures training in Virtual Reality.

As relayed by [49], Adams [9] suggested positive effects of Virtual Reality for technical procedures training performances. They noticed that VR trainees learned faster and made few errors. In the military field, [25] also demonstrated significant learning improvements for Virtual Reality on naval technical procedures. Finally, the literature also shows that Virtual Reality simulators have proven effectiveness on the skills transfer [24, 32, 43, 44, 46].

Simulation's fidelity level tend towards training efficacy. [11] showed that, for driving learning simulators, the training transfer effectiveness varied with simulation fidelity. Results showed that learners faced fewer accidents when they used to practice before on the highest fidelity simulator. Haptics is one of the main components of Virtual Reality simulators fidelity level. [27] defined passive haptics as "a technique that incorporates passive physical objects into virtual environments to physically simulate the virtual objects". Several works show that including passive haptics in virtual environments can increase the sense of presence [27, 41], which enhance skills acquisition and transfer of knowledge [39]. As noticed by [42], passive haptics lack can affect tasks that an operator has to perform. He studied the feasibility of a marksmanship training system in Virtual Reality, by using commercial off-the-shelf systems. He justifies the use of off-the-shelf systems by underlining that, for development of military simulators, initial and operational costs are lower than custom-made systems. That is why he noticed that many works tried to employ only off-the-shelf systems for training solutions [12, 13]. Nevertheless, he did not produce usability, training effectiveness and training transfer studies, which are part of future works.

Learner's motivation is one of the main factors of learning efficiency [52], and it appears that Virtual Reality can foster it largely [18, 48]. [49] mention that *VR enhances trainees' intrinsic motivation* and bind it to freedom degrees and "high-tech" nature of this technology. As relayed by [22], intrinsic motivation involves people doing an activity because they find it interesting and derive spontaneous satisfaction from the activity itself.

Some works also tended to show positive effects of Virtual Reality on learners' self-confidence. During their experiments on VR effectiveness for emergency procedures learning, [53] noticed that subjects expressed their increased confidence in performing their tasks.

In order to be efficient, a training tool must be as accessible and comfortable as possible. One of the main issue important to consider is cybersickness, which can bring users to reject the tool or miss learning objectives. As explained by many authors [21, 33], sensory conflicts are mainly behind cybersickness issues. Visually moving through the virtual environment while staying stationary in the real world is one of the main causes. That is why some works of the literature opted rather for teleport locomotion techniques [14, 15].

We produced a Virtual Reality prototype based on the *EAO FELIN*



Figure 1: FELIN soldier equipped with the IR sight on the FAMAS rifle [3]

software functionalities, and we experimented it on infantrymen learners. To corroborate the already known positive effects of Virtual Reality on training, we want to confirm if it could be applied for rifle maintenance procedure training. We wanted to verify that, by analyzing committed errors and other factors like intrinsic motivation to repeat the training task, mistake awareness and self-confidence. Finally, we wanted to determine how simulation fidelity could impact learning efficiency and if passive haptics from off-the-shelf systems are relevant for rifle maintenance procedure training, as suggested for other use cases from the literature.

3 FELIN

3.1 The FELIN combat system

The FELIN system, developed by Safran Electronics Defense [7], improves a modified FAMAS rifle with a set of clothes, pouches, body armors along with electronic, optic and optronic devices.

One of these devices is the IR sight, an infrared vision sight device fixed over the infantryman's FAMAS rifle (Fig. 1). It contains optronic components allowing night and un-camouflaged vision. When a soldier looks by the sight eyepiece, he can display an interface and navigate within some specific menus. To interact with this interface, the infantryman uses a remote control fixed on the rifle's handle (Fig. 1, soldier's left hand). He can also use other buttons, placed on each side of the IR sight (on the side of the sight on Fig. 1).

3.2 FELIN IR sight's calibration procedure training

To guarantee the operational maintenance of his weapon, the soldier must be efficient in the calibration of his IR sight. He realizes this specific procedure on the shooting range. By navigating within the sight interface, the soldier has to merge the sight and the barrel axis. He controls their alignment by shooting five cartridges on a target. He moves to the target and estimates the Medium Impact Point (PMI) (Fig. 2, left), by using the position of the five impacts. Then, he must enter the correction on the sight interface and repeat the process as many times as needed. Two different corrections are needed: height correction (vertical shift between sight and barrel axis) and direction correction (horizontal shift between rifle and barrel axis).

3.3 EAO FELIN software

Today, the French infantryman train to the FELIN combat system by using a software called *EAO FELIN*, developed by Capgemini. This software includes a specific module for the IR sight calibration training. To practice the procedure, learners use a 2D interface showing the remote control interface, the sight's buttons and the trigger of the rifle (Fig. 2, right).

Currently, for pedagogical and safety reasons, soldiers practice on the software until making no mistakes. When they succeed, they receive a procedure certificate and are allowed to experiment the real situation on the shooting range.

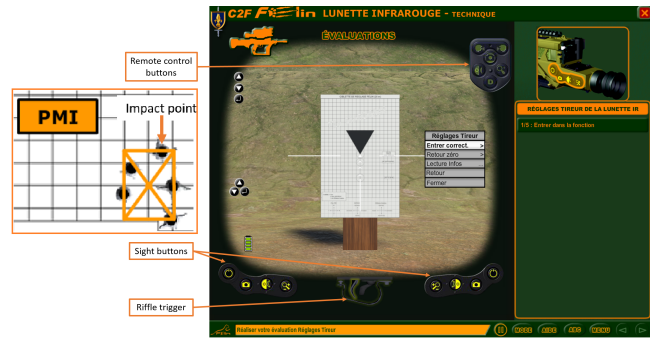


Figure 2: Medium Impact point (PMI)(left) and the sight calibration training module on the *EAO FELIN* software(right)

4 VIRTUAL REALITY PROTOTYPE

4.1 Principle

The VR prototype has been developed to assess the potential learning improvements that Virtual Reality can provide on the IR sight calibration procedure. The goal is to focus on the training procedure only, so we ported the *EAO FELIN* software procedure on Virtual Reality (Fig. 3).

This procedure requires the soldier to consecutively take the shooting posture, initialize the sight through its interface, shoot five cartridges, and go on the target to control the PMI. Finally, he must return to the shooting range and validate the procedure if the PMI is the targeted point. Otherwise, he has to enter the correction values through the sight interface and repeat the procedure until the PMI matches the targeted point.

4.2 Implementation

We developed this prototype with Unity 3D [1], and we used an HTC Vive headset [8]. The virtual environment was displayed in stereoscopy, and the sight view was displayed on the right eye when the user approached his right eye to the lens filter (Fig. 4, left).

We exploited the possibilities of Virtual Reality by extending some procedure's steps of the software. We modeled a shooting range with a target. Trainees have to take the prone shooting posture and move between the rifle and the target position. The following steps were implicit or facilitated on the 2D software:

- *Take the shooting posture.* This step is absent in the 2D software, but we gave the users the possibility to take the prone shooting position in Virtual Reality. For the first version, we used an off-the-shelf rifle's support [5] for the Vive controllers. The controllers' fastenings were magnetized on the structure, and we adjusted their position to match with the FAMAS handle positions (Fig. 5, left). We also fixed an off-the-shelf bipod [2]. This material combination was intended to allow the trainee to take the prone shooting position. The goal was to determine if passive haptics of these off-the-shelf props could bring a suitable level of fidelity.
- *Shoot 5 cartridges.* While on the 2D software the user had to click only once on the trigger button, we let the user start each shot by a pressure on the virtual rifle's trigger. We assigned the trigger of one controller to the trigger of the rifle. The shooting behavior then became more faithful in Virtual Reality.
- *Go to the target.* After the shots, the 2D software automatically zoomed on it. We extended this step in Virtual Reality by allowing the user to teleport to the target. A teleportation zone was placed on the side of the shooting range (Fig. 4, right), to allow the learner to go on the target. When the user entered the zone, he teleported instantly in front of the target. For going back to the rifle, we placed a teleportation zone near the target.

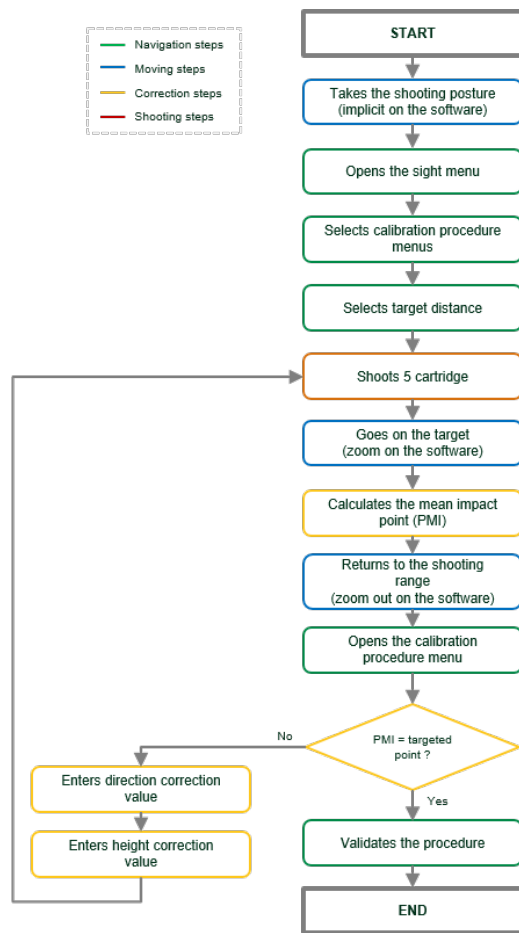


Figure 3: Training procedure, applied to the 2D software and the VR prototype

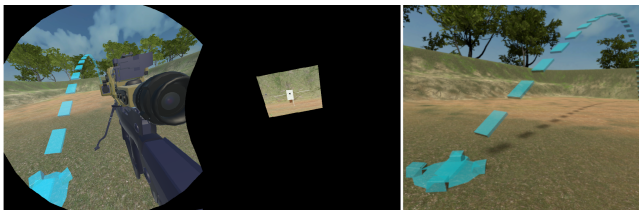


Figure 4: Stereoscopic view with the sight view on the right eye (left), teleport zone (right).

- *Measure the PMI.* On the 2D software, the PMI is displayed on the target for 5 seconds after the zoom. In Virtual Reality, it is also displayed on the target, but there were no time restrictions.

To control the sight's interface, the trackpad of the second controller was assigned to the remote control. In order to be able to find the buttons position when equipped with the headset, an adhesive mask taking the shape of the remote control buttons was placed over the trackpad (Fig. 5, right). We divided the trackpad zone into four areas, and each one matched a remote control button.

The VR prototype used the same rules present in the 2D software for each procedure step. If the user's action is unexpected, it is impossible to navigate in the wrong menus of the interface, to shoot and to move to the target. If the learner tries to do an unexpected action, the system indicates that he has made a mistake by playing a sound, just like in the *EAO FELIN* software. We traced these mistakes and all the actions done in a log file. We classified mistakes in three different types:



Figure 5: VR rifle support vs. FAMAS rifle (left) and adhesive mask added on the Vive controller (right)

- **Procedure mistakes:** the learner tried to go to a wrong menu or make actions at the wrong time (shooting, moving to the target).
- **Correction mistakes:** the learner did not enter the right correction value, or he mistook the height or direction corrections.
- **Button mistakes:** the learner used the wrong button or did not press it for enough time if a long push was needed.

As the goal was not to train on shooting, the bullet impacts were randomly computed. There was a tolerance concerning the shooting position, as well as aiming and shooting the target.

5 EXPERIMENTS

During the experiments, we compared the two learning methods through to two angles. First, we conducted user tests on the VR prototype. Afterward, we wanted to see if Virtual Reality showed potential benefits for this specific use case. We wanted to observe if allowing the learner to live an experience closer to the real one would impact the learning efficiency of the procedure.

5.1 User tests

For user tests, we wanted to see if the users' interactions with the different elements were pleasant, enjoyable and realistic. We focused ourselves particularly on the rifle, the sight, the remote control, the menus, and the moving method. Finally, we wanted to see if there was a difference between the feelings of the learners and those of the instructors. We believe that comparing these results could let us spot potential lacks only identifiable by experienced users.

5.2 Learning improvement hypothesis

For this use case, we formulated the following hypotheses:

- H1** Learners will be more intrinsically motivated to repeat the training task when using Virtual Reality rather than the 2D software.
- H2** Learners will make fewer mistakes after the use of Virtual Reality.
- H3** Learners will need fewer replays until making no mistakes on the 2D software, if they used the Virtual Reality prototype on the first try.
- H4** Learners will have a better awareness of their mistakes in Virtual Reality.
- H5** Learners will have a better self-confidence after the use of Virtual Reality.

5.3 Experimentation process

5.3.1 Groups

To check these hypotheses, we opted for the experimentation process shown in Fig. 6. At present, after following a set of lessons about the *FELIN* system and a summary on how to handle the equipment, learners practice on the *EAO FELIN* software until they make no mistakes. So, we chose to split learners into two groups:

- a control group, on the 2D software learning method, iterating on the 2D software until making no mistakes. Before using the 2D software, we gave them a short set of instructions.

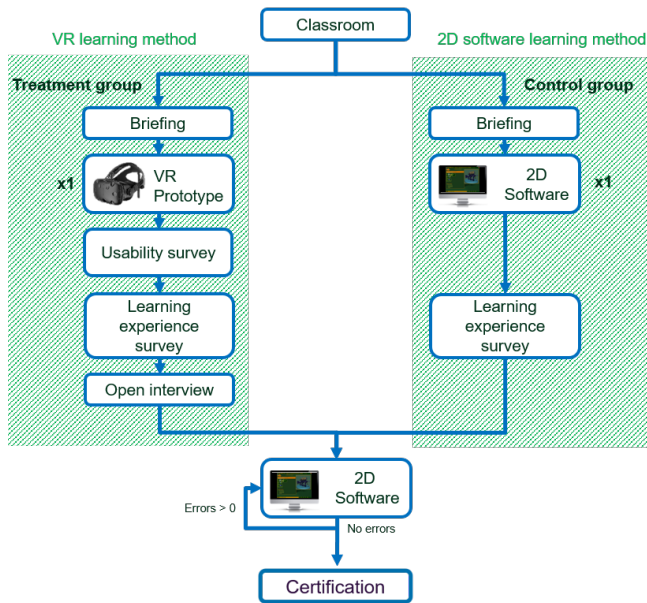


Figure 6: Experimentation process

Table 1: User tests questionnaire

Questions (Likert [Totally disagree(1)-Totally agree(5)])
1- Handling rifle was easy
2- Handling rifle was unpleasant
3- Handling rifle was realistic
4- Finding the way to look through the sight was easy
5- Looking through the sight was unpleasant
6- Looking through the sight was realistic
7- Finding remote control buttons was easy
8- Using the remote control was easy
9- Using the remote control was unpleasant
10- Using the remote control was realistic
11- Menus were enough visible
12- Finding the way to go on the target was easy
13- Going to the target was unpleasant

- a treatment group, on the VR learning method, making their first iteration on the VR prototype before iterating on the *EAO FELIN* software until making no mistakes. Before using the VR prototype, they also were given a short set of instructions that presented the prototype and its functioning.

In the treatment group, learners used the VR prototype independently from one another, only accompanied by an instructor and an experimenter. The aim was for learners not to see the procedure in VR, to avoid biases on the experiment. When they used the 2D software, learners of each group were also in the same computer room, equipped with headphones.

5.3.2 User tests

After prototype use, learners from the treatment group were invited to answer a set of Likert [35] scale questions (Table 1). The purpose of this survey was to evaluate the VR system, focusing on different elements like the rifle, the sight, the remote control, the menus, as well as the moving method and its metaphor. We included the IPQpresence survey [6, 45] too, which allows assessing presence through three subscales: Spatial Presence (sense of being physically present), Involvement (attention on the virtual environment, involvement experienced) and Experienced Realism (subjective experience of realism).

Finally, after answering these surveys, learners from the treatment group were encouraged to express feedback about their experience, during an open interview. They lasted about five minutes and started by the same initial question: "Could you express any feedback you

feel relevant about the experience?". The goal was to let appear elements which were not detectable by the surveys. We recorded each point raised by the subject and synthesized them when we processed the results manually.

To gather feedback from experienced users, we also tested the VR prototype on instructors. We gave them the same survey followed by an open interview.

5.3.3 Learning benefits experiments

After their first iteration, both groups also had to answer a Likert survey which focused on their learning experience. The questions leaned towards specific points we wanted to analyze: user pleasure and interest to use the learning medium, his intrinsic motivation to repeat the training task and his self-confidence. Two Likert scales focused on the pleasure and interest to use the learning medium ("Performing the procedure was felt like a pleasure.", "The used medium was attractive") and another one on the intrinsic motivation to repeat the training task ("Performing again the procedure would be a pleasure."). We also added a question where learners had to estimate approximately the number of mistakes they made. This information was used to analyze the awareness of their mistakes by comparing it to the log file.

5.3.4 Results exploitation

We calculate a score (arithmetic mean) from answers to questions for each element analyzed. For the negative questions (e.g., "The element use was unpleasant"), the value taken for the mean was the answer value subtracted from the maximum value (5).

5.4 Subjects

We led experimentations on 76 learners, lieutenants of the Draguignan Infantry School [4], from 6 different squads. 22 learners composed the treatment group and the 54 others the control group. We did not use randomization, but instructors were asked to select learners randomly for the treatment group.

The prototype implemented was only adapted for right-dominant eye users, and it was the unique constraint required before the treatment group random selection.

The average age of subjects was 27(± 4) years old (28(± 7) for the treatment group, 24(± 6) for the control group), and they were 99% male. In the treatment group, 74% of them had never used a Virtual Reality headset before.

For user tests, we also experimented it on 6 instructors of the military school. They were all male, their average age was 41(± 5) years old, and 67% of them had never used a Virtual Reality headset before.

5.5 Time and logistical constraints

We made an effort to engage end users for this study. Using convenient subjects would not be a relevant way of executing it in this domain. Engaging end users in studies is hard to achieve, and we are grateful that the French Army gave us this opportunity. Nevertheless, we faced up to time and logistical issues. Our study was integrated to the military school real learning session and was limited by time and availability of learners and instructors.

That is why it was not possible for us to isolate subjects using the 2D software. However, instructors supervised and kept a close watch on them, to avoid exchanges.

It also made impossible to have two groups of the same size. Nevertheless, we produced statistical analysis aimed to notice significant results with two samples of different sizes.

6 RESULTS

6.1 Methodology

We performed statistical analysis for each result, by comparing the results of each group as two independent samples.

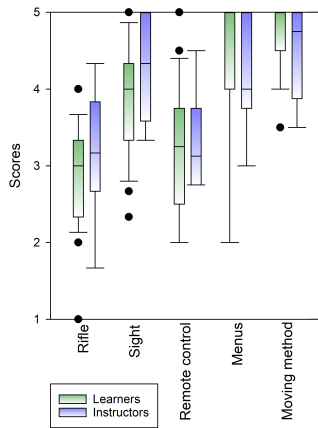


Figure 7: VR prototype assessment results

For data from the Likert scale values, we proceeded with the Mann-Whitney U test [38] to see if there was a statistically significant difference between our treatment and control groups. In the same way, for other data like the number of mistakes and iterations, we check if they had a normal distribution (Shapiro-Wilk test [47]) and a homogeneous variance (Brown-Forsythe test [17]). If it was the case, we proceeded with the Student's t-test [51]. If not, we also used the Mann-Whitney U test.

Finally, for the number of mistakes reduction on each replay, we used the same test to check normal distribution (Shapiro-Wilk), and if it succeeds, we also used the Student's t-test. If not, we used the Wilcoxon signed-rank test [56].

6.2 First iteration results

The first iteration let us conduct user tests on the VR prototype and analyze learning impacts feelings (pleasure and interest to use the learning medium, intrinsic motivation to repeat the training task and self-confidence feelings).

6.2.1 User tests

We compared the results of the prototype assessment from the surveys answered by learners and instructors (Fig. 7).

The rifle (2.94 ± 0.68) and the remote control (3.18 ± 0.83) were moderately rated. Three questions were related to the rifle, and four were related to the remote control. As shown in Table 3, the lack of fidelity of passive haptics from the rifle seemed to make its handling quite unpleasant. For the remote control, users felt significant difficulties to find the buttons. This effect seemed to have affected them by making its handling quite unpleasant too. Open interviews tended to confirm these results. A significant number of users freely indicated they were uncomfortable in finding the remote buttons (65%), and most of them declared they wanted to see their hands for placing their thumb correctly on buttons (60%). Passive haptics of the rifle were also pointed out for lacks of fidelity. Many learners underlined that the gun weight was too light (48%) and then experienced some difficulties. Another criticized aspect was the magnetic system of the support. Infantrymen tend to strongly grab the handles of the rifle on their shooting position, and with this system, the controllers slightly moved. Finally, some users confessed being bothered by the fidelity lacks of the trigger (22%) and the absence of recoil after a shot (13%). Learners expressed different opinions about these lacks. Several of them expressed the need for "recovering real sensations" during the experience, and some others mentioned that the absence of recoil let them be "more focused on the calibration". All these fidelity aspects have also been brought up by instructors during their open interview.

Subjects rated well the General Presence (3.87 ± 1.15) and Spatial Presence (3.72 ± 0.64) (Fig. 8). However, involvement (3.06 ± 0.60) and Experience realism (2.89 ± 0.43) received average scores. For

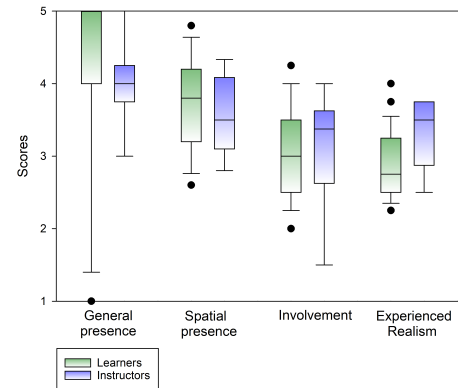


Figure 8: VR prototype - presence results (IPQ)

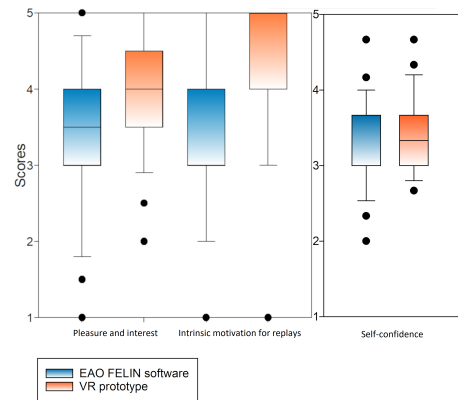


Figure 9: Learning experience questionnaire results

each scale of presence, there was not a statistically significant difference between learners and instructors (Table 2, PRDIFF).

The sight (3.93 ± 0.71), the menus (4.17 ± 1.05) and the moving method (4.76 ± 0.44) were rated satisfactorily (Fig. 7).

For each item (rifle elements, satisfaction, presence), there was not a statistically significant difference between learners and instructors (Table 2, UTDIFF).

6.2.2 Learning benefit results

Scores of pleasure and interest to use the learning medium and intrinsic motivation to repeat the task, showed a statistically significant difference (Table 2, ISDIFF, MSDIFF) between treatment and control groups (Fig. 9). Users who used the VR prototype expressed more pleasure and interest (Med: 4.00) than the *EAO FELIN* group (Med: 3.5). They also seemed to feel more intrinsically motivated to repeat the task (Med: 4.00) than the *EAO FELIN* group (Med: 4.00).

Nevertheless, there was no statistically significant difference for the learner's self-confidence (Table 2, SCDIFF).

6.2.3 Error awareness

In order to analyze the learners' error awareness, we calculated the approximation errors of the mistake estimations from both groups (Fig. 12, right). We compute it from their number of mistakes made (saved in the log files) and their estimated amount (answered in the learning experience questionnaire). These results did not show a statistically significant difference between the control and the treatment group (Table 2, EEDIFF).

6.3 Replays results

Results from replays let us analyze the impacts of the VR learning method on mistakes, number of needed replays and time spent on them.

Table 2: Statistic tests results

		Test	Results	Significant?	Cohen's d
First iteration results					
UTDIFF	User tests differences learners-instructors	Mann-Whitney U test	$p > 0.05$	Not significant	
PRDIFF	Presence differences learners-instructors	Mann-Whitney U test	$p > 0.05$	Not significant	
ISDIFF	Medium pleasure and interest scores	Mann-Whitney U	$U = 426.5, z = -2.59, p = 0.018$	Significant	0.29
MSDIFF	Intrinsic motivation scores	Mann-Whitney U	$U = 421, z = -2.65, p = 0.019$	Significant	0.30
SCDIFF	Learner's self-confidence	Mann-Whitney U	$U = 579.5, p > 0.05$	Not significant	
EEDIFF	Approximation of mistake estimations	Shapiro-Wilk Mann-Whitney U	Normality test failed $U = 556.5, p > 0.05$	Not significant	
Replays results					
PMSRED	Procedure mistakes-Soft. errors reduction	Wilcoxon	$T = 16, z = -4.342, p < 0.001$	Significant	0.69
PMVRED	Procedure mistakes-VR proto. errors reduction	Wilcoxon	$T = 0, z = -3.461, p < 0.001$	Significant	0.89
PMCOMP	Procedure mistakes-Soft./proto. comparison	Shapiro-Wilk Mann-Whitney U	Normality test failed $U = 246.5, z = 1.22, p = 0.024$	Significant	0.16
CMSRED	Correction mistakes-Soft. errors reduction:	Wilcoxon	$T = 68.5, p > 0.05$	Not significant	
CMVRED	Correction mistakes-VR proto. errors reduction	Wilcoxon	$T = 0, z = -2.414, p = 0.016$	Significant	0.60
CMCOMP	Correction mistakes-Soft./proto. comparison	Shapiro-Wilk Mann-Whitney U test	Normality test failed $U = 229.5, z = 0.92, p = 0.011$	Significant	0.12
BMCOMP	Button mistakes-Soft./proto. comparison	Shapiro-Wilk Mann-Whitney U	Normality test failed $U = 292.5, z = 3.45, p = 0.033$	Significant	0.46
NBITER	Number of iterations needed	Shapiro-Wilk Mann-Whitney U test	Normality test failed $U = 45, z = -5.62, p < 0.001$	Significant	0.67

Table 3: Rifle and remote control questionnaire results

Questions (Likert [1-5])	Rifle	
	Learner means	Instructor means
Handling it was easy	3.17 (± 0.87)	3.67 (± 0.47)
Handling it was disagreeable	2.91 (± 1.17)	3.33 (± 1.11)
Handling it was realistic	2.64 (± 0.88)	3.17 (± 1.34)
Questions (Likert [1-5])	Remote control	
	Learner means	Instructor means
Finding buttons was easy	2.48 (± 1.25)	2.83 (± 1.07)
Finding way to use it was easy	3.30 (± 1.04)	4.00 (± 0.82)
Using it was disagreeable	2.52 (± 1.10)	3.00 (± 1.00)
Using it was realistic	3.42 (± 1.06)	3.33 (± 0.94)

We excluded 6 subjects of the treatment group from the results. 4 of them were faced with time issues and did not have the time to finish the 2D software iterations. The 2 others faced technical issues with the 2D software, which could have impacted their results. We excluded 14 subjects of the control group from the results. 11 of them was faced time issues and did not have the time to finish the 2D software iterations. During the replays, 2 others did not respect the instructions and restarted the software when they were making errors. Finally, the last one faced technical issues with the 2D software during replays, which could have impacted his results.

Consequently, the final population studied for the replay results was 51 subjects (15 in the treatment group and 36 in the control group).

For the mistake estimation, we excluded 1 subject from the treatment group and 4 subjects from the control group because they did not answer the mistake estimation question. The final population studied for the mistake estimation was 46 subjects (14 in the treatment group and 32 in the control group).

6.3.1 Mistakes

Experiments showed interesting results about mistakes between the treatment and the control group (Fig. 10). To exclude eventual bias, we only compared mistakes which were possible in both environments (VR prototype and *EAO FELIN* software). For the treatment group, we rejected mistakes which were due to a manipulation error (e.g., accidental push on a button while catching the rifle).

There was a statistically significant difference for the mistakes made on replays. Learners who used the VR learning method did not make any procedure and correction mistakes, while 2D software method users continued to make some (Table 2, PMSRED, PMVRED, PMCOMP, CMVRED, CMCOMP). Nevertheless, for

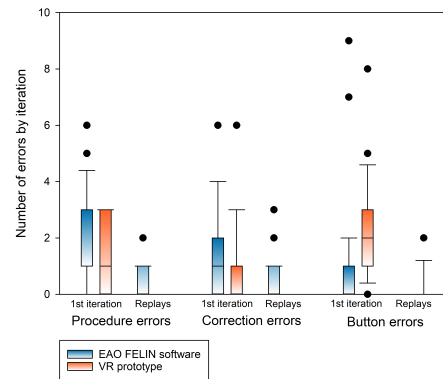


Figure 10: Mistakes made on the first iteration and mean of mistakes made on each iteration of the replay

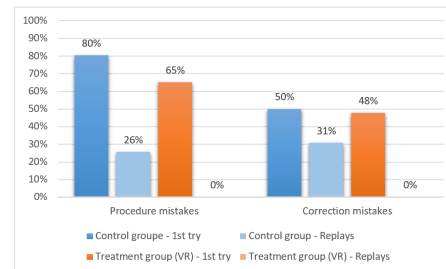


Figure 11: Procedure and correction percentage of mistakes

the correction mistakes of 2D software learners, the difference is not statistically significant (Table 2, CMSRED). These effects are more flagrant when we analyze the percentage of learners making mistakes (Fig. 11).

The opposite effect appeared for button mistakes. On their first iterations, the treatment group made more mistakes (Med: 0) than the control group (Med: 0) (Fig. 10). This result was statistically significant (Table 2, BMCOMP). Another interesting result came from the fact that users who made their first iteration on the 2D software did not make any button mistakes when VR learners continued to make some.

6.3.2 Replays needed

VR learners needed fewer replays until making no mistakes than the 2D software ones (Fig. 12, left). Learners from the treatment

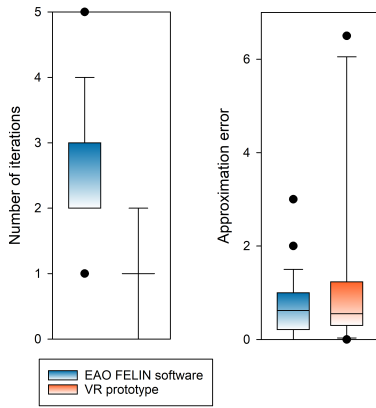


Figure 12: Replays needed until making no more mistakes (left) and mistakes estimations' approximation errors (right)

group needed significantly fewer iterations (Med: 1) until making no mistakes than learners from the control group (Med: 3) (Table 2, NBITER). We notice that there is a significant important effect (Cohen's $d = 0.67$) on the number of needed replays.

It is interesting to notice that 2 learners from the treatment group made no mistakes on their first try on the VR prototype. However, they made an additional iteration on the *EAO FELIN* software to validate their certification.

7 DISCUSSIONS

7.1 Results interpretation

7.1.1 Fidelity lacks

User tests underlined fidelity lacks from passive haptics of the rifle and the remote control. Instructors' results and open interviews tended to confirm these results.

These fidelity lacks seemed to have mixed effects on our results. That is why we suppose it is relevant to tend towards a high level of fidelity when the training purpose concerns a procedure for a device where its manipulation requires a high level of sensorial focus.

We supposed that the users' desire to visualize their hands in the virtual environment was due to the touch-sensitive differences between the trackpad-adhesive assembly and the real remote buttons. Real remote buttons are indeed large and accompanied by a tactile cue to easily help to find their positions.

Finally, results underlined a good assessment for menu visibility and ergonomics. The sight functioning was well rated too, confirming our implementation choice to render the sight's camera on the dominant-eye when the user approaches it, close to the lens filter.

7.1.2 Sense of presence

We analyzed the sense of presence in order to help to detect potential lacks on the VR prototype.

As noticed in the Results section, Involvement and Experienced realism were moderately rated. It is highly likely that they could have been impacted by the rifle fidelity lacks.

Finally, we could have supposed that the presence of the instructor and the experimenter in the same room could have impacted the Involvement scale of presence, but that seems doubtful because they paid attention to be particularly discreet and act only in last resort.

7.1.3 Users' expertise

As noticed previously, there was no statistically significant differences between learners and instructors on presence and system assessments. These results let us suppose that the prototype assessment is probably not affected by the users' expertise and that the system did not deviate from actual field conditions. Nevertheless, as the sample of instructors was thin, we cannot draw any conclusions about it. However, we believe it is important to check if there is a

Table 4: Hypothesis results

#	Hypothesis	Results
H1	More intrinsically motivated	Validated
H2	Fewer mistakes on replays	Partially validated
H3	Less replays needed	Validated
H4	Better awareness of mistakes	No significant results
H5	Better self-confidence	No significant results

significant difference between learners and their instructors to be able to detect potential bias linked to the users' expertise.

7.1.4 Learning improvements

As summarized in Table 4, experiments allowed us to validate hypotheses **H1** and **H3**. Hypothesis **H2** was partially validated, due to fidelity lack issues. Hypotheses **H4** and **H5** did not show significant results and require more improvements and experiments.

H1: Learners will be more intrinsically motivated to repeat the training task when using Virtual Reality rather than the 2D software. Results showed that learners felt more pleasure and interest with the VR prototype. These effects implied positive impact on the intrinsic motivation to repeat the training task. It validated the assumption that living experience in Virtual Reality for a military device calibration procedure should enhance learners' intrinsic motivation to repeat the training task. Although this result is significant, the effect size is low (Cohen's $d = 0.30$).

H2: Learners will make fewer mistakes after the use of Virtual Reality. One of the more interesting improvements revealed by the results came from the analysis of mistakes made after the use of the VR learning method. After its use, learners using the *EAO FELIN* software made no procedure or correction mistakes. On the other hand, the 2D software learners continued to make this kind of errors on the following attempts. These results showed that living experience within an immersive simulated situation facilitates the activation of cognitive levers, necessary to assimilate the procedure.

However, these results only partially validate the hypothesis because we noticed the opposite effect on button mistakes. We suppose that the fidelity lacks previously discussed most probably impact these results.

H3: Learners will need fewer replays until making no mistakes if they used the VR prototype on the first try. Learners in VR needed fewer iterations than software users. This results is strongly linked to the fact that learners made fewer mistakes after the use of Virtual Reality (**H2**).

H4: Learners will have a better awareness of their mistakes in Virtual Reality. The mistakes estimation did not show significant results. We observe an important data scattering for the approximation of errors of the VR learners about the estimation of their number of mistakes. It brought us to suppose that these results could highly come from the fidelity of the remote control.

H5: Learners will have a better self-confidence after the use of Virtual Reality. Results about self-confidence did not show significant results. We suppose that it could be due to the procedure simplicity.

7.2 Observations

Experiments helped to highlight the limits of the *EAO FELIN* software. Indeed, it was interesting to notice that Virtual Reality helped to detect errors not considered on the 2D software. For example, some of them tried to navigate in menus instead of moving to the target to check the PMI. In the 2D software, this step occurred automatically by zooming on the target after the shot.

We also observed that in Virtual Reality, even if learners were informed that shooting accuracy was not necessary for the training situation, they still paid much attention to aim the target. They used to stop their breath at shooting time and control their pressure on the trigger. We interpret this fact by the sense of presence provided by

the VR prototype, which allowed learners to perform their procedure conscientiously.

7.3 Limits

We identified two potential limits for our results. The first one is the simplicity of the procedure. Indeed, the IR sight calibration procedure is quite simple. We suppose that some results which were not statistically significant could be more pronounced for other procedures, depending on their difficulty.

The second potential limit comes from the population of subjects on which we did experiments. They all were lieutenant trainees, which mean that they had just recently graduated from the officer school. We suppose that their educational level could have affected some of our results which were not statistically not significant.

8 FUTURE WORKS

8.1 New prototype versions

The most important improvement axes of our future works must consist in reducing the passive haptic gaps noticed, and comparing the effects on the learners' behaviors and performances. Some works of the literature [30, 31, 50] suggest that *haptic feedback in an early training phase may improve the trainee's performance, by enhancing the trainee's sensory perception capabilities and thus facilitating the transfer of skill* [55]. One of the possible improvements could consist of using demilitarized equipment as passive haptic systems, ideally modified with a pneumatic system. Comparing the learners' concentration and performance, by varying on these different fidelity levels, could be interesting. In this purpose, we plan to lean on frameworks from literature, like AFFECT [40].

Some literature works suggest that the effects of body visibility in the virtual world can create an adaptation time for the learners, which can affect significantly positively their learning performance [29]. At this step of the prototype development, the user cannot see his hands in the virtual environment. This is why we plan to equip some users with data gloves on a high fidelity version of the prototype and then analyze if allowing learners to see their hands impacts their efficiency and their committed mistakes.

To focus on the learning procedure, we chose to generate shooting impacts randomly on the target and to indicate the PMI and the correction values required on the target. We think that making a more faithful prototype version and compare it to a new 2D software version could be interesting. In this version, the aimed point could be the shot one, and the PMI and the correction values could be deliberately absent. This faithfulness improvement could allow analyzing if Virtual Reality impacts sight behavior and value estimations.

Instructors suggested to enlarge the system to assess some good shooting practices during the experience (breath control, pressure levels on the trigger). It could be a good prospect as part of future works. We could then analyze if learners are as focused and attentive to the procedure when they also have in mind their good shooting practices.

Finally, we also plan to develop a left dominant-eye prototype version in order to have more representative samples of subjects.

8.2 Additional studies

Once the fidelity issues resolved, we plan to analyze usability by carrying out experiments based on SUS method [16].

We also plan to verify if fidelity issues impacted button errors results. Moreover, 2D software users did not make any errors of this kind on replays. We want to analyze if it is also the case on the VR prototype.

We want to see if fidelity issues affected errors estimation results too. We plan to distinguish the different kinds of mistakes in this estimation. If there is a significant difference, we plan to analyze if results changes according to the type of mistake.

We did not measure the sense of presence for the control group. In future works, we plan to measure it and compare it with the treatment group results. We also want to compare presence values with new values after improving the fidelity levels of the rifle.

During open interviews, some learners noticed a "*sensation of calm*" through the VR procedure execution. We want to focus on this potential effect, in order to try to identify its causes and analyze how it could be an interesting learning lever for training through Virtual Reality.

To confirm motivation impact noticed and to clarify the size of its effect, we should adapt and apply more advanced motivation assessment methods from the literature [54]. The fact that 2D software learners were performing in the same room at the same time could also have impacted the motivation. More experiments, with isolated subjects, could help disprove this hypothesis. Finally, because the training purpose took place in a commanded military certification context, focusing on extrinsic motivation potential effects [22] could be interesting.

We supposed the procedure simplicity impacted self-confidence results. To confirm this assumption, we plan to analyze it on more complex procedures. We will also explore literature and apply more advanced self-confidence assessment methods. Finally, we want to add preliminaries questions to evaluate users' initial self-confidence.

The literature shows that using repetitive VR training can improve the learning curve before practicing on real situations [10, 20, 44]. In our future works, we plan to let each group iterate on the same medium (software or VR prototype), then analyze the total errors made, time spent and iterations needed to assess if it implies training time gains and effectiveness.

We suppose that the high educational level of subjects could have affected some of our results which were not statistically not significant. We want to experiment the VR prototype on learners having a lower military rank and compare our results.

Finally, in our experiments, we analyzed the learning improvements by comparing the performances of trainees through replays on the software. We presume that we could also notice some improvements when VR learners perform their first real procedure on the shooting range. We strongly suppose that VR learners will spend less time making their first calibration procedure in the real situation. Experiment it could be an interesting way to analyze the learning transfers effects of Virtual reality for this use case. We plan to use our current results in order to obtain the opportunity to access to learners during their first procedures on the real shooting range and carry out a training transfer study.

9 CONCLUSIONS

In this paper, we presented an ad-hoc study of a Virtual Reality prototype used for training on a military sight calibration procedure. We experimented it on real infantry learners from an infantry school.

Results showed that for this kind of procedures, Virtual Reality could improve learners' intrinsic motivation to replay the training task. The prototype also showed interesting results in learning efficiency. VR learners made no errors when 2D software learners continued to make some. Experiments let us identify some fidelity lacks on the prototype, needing improvements. Finally, results also helped identify some procedure mistakes which are currently not detectable by the traditional software.

ACKNOWLEDGMENTS

The authors wish to strongly thank the Human Resources Department of the French Land Army (DRHAT) and the Infantry School of Draguignan (EMD) for allowing us to experiment the prototype on real infantryman learners and also publishing this results.

REFERENCES

- [1] <http://unity3d.com/>.
- [2] <http://www.amazon.fr/gp/product/B00M9RF0ZM/>.

- [3] <http://www.defense.gouv.fr/terre/equipements/materiels-generiques/equipement/felin-fantassin-a-equipement-et-liaisons-integres>.
- [4] <http://www.emd.terre.defense.gouv.fr/>.
- [5] <http://www.etsy.com/listing/491103695/htc-vive-magnetic-dual-controller-rifle>.
- [6] <http://www.igroup.org/pq/ipq/index.php>.
- [7] <http://www.safran-electronics-defense.com/>.
- [8] <http://www.vive.com/>.
- [9] N. Adams. A study of the effectiveness of using virtual reality to orient line workers in a manufacturing environment. *Motorola University, unpublished dissertation*, 1996.
- [10] R. Aggarwal, T. P. Grantcharov, J. R. Eriksen, D. Blirup, V. B. Kristiansen, P. Funch-Jensen, and A. Darzi. An evidence-based virtual reality training program for novice laparoscopic surgeons. *Annals of surgery*, 244(2):310, 2006.
- [11] R. W. Allen, G. D. Park, M. L. Cook, and D. Fiorentino. The effect of driving simulator fidelity on training effectiveness. *DSC 2007 North America*, 2007.
- [12] N. L. Arthur. Proof-of-concept part-task trainer to enhance situation awareness for instrument approach procedures in aviation domain. Technical report, Naval Postgraduate School Monterey United States, 2017.
- [13] J. T. Attig. *Proof-of-concept part task trainer for close air support procedures*. PhD thesis, Monterey, California: Naval Postgraduate School, 2016.
- [14] D. A. Bowman, D. Koller, and L. F. Hodges. Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. In *Virtual Reality Annual International Symposium, 1997., IEEE 1997*, pp. 45–52. IEEE, 1997.
- [15] E. Bozgeyikli, A. Raij, S. Katkooi, and R. Dubey. Point & teleport locomotion technique for virtual reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, pp. 205–216. ACM, 2016.
- [16] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.
- [17] M. B. Brown and A. B. Forsythe. Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346):364–367, 1974.
- [18] G. C. Burdea and P. Coiffet. *Virtual reality technology*. John Wiley & Sons, 2003.
- [19] A. Clavarelli, W. L. Platte, and J. J. Powers. Teaching and assessing complex skills in simulation with application to rifle marksmanship training. Technical report, NAVAL POSTGRADUATE SCHOOL MONTEREY CA, 2009.
- [20] P.-J. Fager and P. von Wowern. The use of haptics in medical applications. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 1(1):36–42, 2004.
- [21] P. Fuchs. *Virtual Reality Headsets-a Theoretical and Pragmatic Approach*. CRC Press, 2017.
- [22] M. Gagné and E. L. Deci. Self-determination theory and work motivation. *Journal of Organizational behavior*, 26(4):331–362, 2005.
- [23] N. Gavish, T. Gutiérrez, S. Webel, J. Rodríguez, M. Peveri, U. Bockholt, and F. Tecchia. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, 23(6):778–798, 2015.
- [24] T. P. Grantcharov, V. Kristiansen, J. Bendix, L. Bardram, J. Rosenberg, and P. Funch-Jensen. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *British Journal of Surgery*, 91(2):146–150, 2004.
- [25] R. T. Hays, D. A. Vincenzi, A. G. Seamon, and S. K. Bradley. Training effectiveness evaluation of the vesub technology demonstration system. Technical report, NAVAL AIR WARFARE CENTER TRAINING SYSTEMS DIV ORLANDO FL, 1998.
- [26] R. W. Hill Jr, J. Gratch, S. Marsella, J. Rickel, W. R. Swartout, and D. R. Traum. Virtual humans in the mission rehearsal exercise system. *Ki*, 17(4):5, 2003.
- [27] B. E. Insko, M. Meehan, M. Whitton, and F. Brooks. *Passive haptics significantly enhances virtual environments*. PhD thesis, University of North Carolina at Chapel Hill, 2001.
- [28] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.
- [29] R. V. Kenyon and M. B. Afenya. Training in virtual and real environments. *Annals of Biomedical Engineering*, 23(4):445, 1995.
- [30] H. K. Kim, D. W. Rattner, and M. A. Srinivasan. The role of simulation fidelity in laparoscopic surgical training. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 1–8. Springer, 2003.
- [31] P. Lamata, E. Gómez, F. Sánchez-Margallo, F. Lamata, F. Del Pozo, and J. Usón. Tissue consistency perception in laparoscopy to define the level of fidelity in virtual reality simulation. *Surgical Endoscopy and Other Interventional Techniques*, 20(9):1368–1375, 2006.
- [32] C. E. Lathan, M. R. Tracey, M. M. Sebrechts, D. M. Clawson, and G. A. Higgins. Using virtual environments as training simulators: Measuring transfer. *Handbook of virtual environments: Design, implementation, and applications*, pp. 403–414, 2002.
- [33] J. J. LaViola Jr. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1):47–56, 2000.
- [34] L. Leroy, P. Fuchs, and G. Moreau. Visual fatigue reduction for immersive stereoscopic displays by disparity, content, and focus-point adapted blur. *IEEE Transactions on Industrial Electronics*, 59(10):3998–4004, 2012.
- [35] R. Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.
- [36] M. Macedonia. Games soldiers play. *IEEE Spectrum*, 39(3):32–37, 2002.
- [37] D. Manca, S. Brambilla, and S. Colombo. Bridging between virtual reality and accident simulation for training of process-industry operators. *Advances in Engineering Software*, 55:1–9, 2013.
- [38] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pp. 50–60, 1947.
- [39] F. Mantovani and G. Castelnovo. The sense of presence in virtual training: enhancing skills acquisition and transfer of knowledge through learning experience in virtual environments. 2003.
- [40] R. P. McMahan and N. S. Herrera. Affect: Altered-fidelity framework for enhancing cognition and training. *Frontiers in ICT*, 3:29, 2016.
- [41] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks Jr. Physiological measures of presence in stressful virtual environments. *Acm transactions on graphics (tog)*, 21(3):645–652, 2002.
- [42] F. Rashid. *Use of VR technology and passive haptics for MANPADS training system*. PhD thesis, Monterey, California: Naval Postgraduate School, 2017.
- [43] R. M. Satava. Virtual reality surgical simulator. *Surgical endoscopy*, 7(3):203–205, 1993.
- [44] M. Schijven, J. Jakimowicz, I. Broeders, and L. Tseng. The eindhoven laparoscopic cholecystectomy training course-improving operating room performance using virtual reality training: results from the first eaes accredited virtual reality trainings curriculum. *Surgical Endoscopy and other Interventional Techniques*, 19(9):1220–1226, 2005.
- [45] T. W. Schubert. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realism. *Zeitschrift für Medienpsychologie*, 15(2):69–71, 2003.
- [46] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. O'Brien, V. K. Bansal, D. K. Andersen, and R. M. Satava. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery*, 236(4):458, 2002.
- [47] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.
- [48] E. M. Sims. Reusable, lifelike virtual humans for mentoring and role-playing. *Computers & Education*, 49(1):75–92, 2007.
- [49] R. Stone. Virtual reality for interactive training: an industrial practitioner’s viewpoint. *International Journal of Human-Computer Studies*, 55(4):699–711, 2001.
- [50] P. Ström, L. Hedman, L. Särnå, A. Kjellin, T. Wredmark, and L. Felländer-Tsai. Early exposure to haptic feedback enhances performance in surgical simulator training: a prospective randomized crossover study in surgical residents. *Surgical endoscopy and other*

- interventional techniques*, 20(9):1383–1388, 2006.
- [51] Student. The probable error of a mean. *Biometrika*, pp. 1–25, 1908.
- [52] A. Sutcliffe. *Multimedia and virtual reality: designing multisensory user interfaces*. Psychology Press, 2003.
- [53] D. L. Tate, L. Sibert, and T. King. Virtual environments for ship-board firefighting training. In *Virtual Reality Annual International Symposium, 1997., IEEE 1997*, pp. 61–68. IEEE, 1997.
- [54] R. J. Vallerand, L. G. Pelletier, M. R. Blais, N. M. Briere, C. Senecal, and E. F. Vallieres. The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and psychological measurement*, 52(4):1003–1017, 1992.
- [55] O. A. Van der Meijden and M. P. Schijven. The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: a current review. *Surgical endoscopy*, 23(6):1180–1190, 2009.
- [56] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945.