



Evaluation of Virtual Reality Training: effectiveness on pilots' learning

Nawel Khenak, Cédric Bach, Stéphane Drouot, Florence Buratto

► To cite this version:

Nawel Khenak, Cédric Bach, Stéphane Drouot, Florence Buratto. Evaluation of Virtual Reality Training: effectiveness on pilots' learning. IHM'23 - 34e Conférence Internationale Francophone sur l'Interaction Humain-Machine, AFIHM; Université de Technologie de Troyes, Apr 2023, Troyes, France. hal-04046414

HAL Id: hal-04046414

<https://hal.science/hal-04046414>

Submitted on 25 Mar 2023

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.

Evaluation of Virtual Reality Training: effectiveness on pilots' learning

Nawel Khenak, Cédric Bach

Human Design Group

Toulouse, France

firstname.lastname@humandesign.group

Stéphane Drouot, Florence Buratto

Airbus Commercial Aircraft

Toulouse, France

firstname.lastname@airbus.com

ABSTRACT¹

The article reports on the evaluation of an immersive Virtual Reality (VR) device for pilots' training. It is compared with the use of a non-immersive 3D training device, which relies on conventional flat displays. Specifically, the evaluation aimed to compare the usability of both devices with a focus on learning effectiveness. Thus, 12 participants performed a learning task using both devices. The data collected included performance learning and satisfaction. In addition, special attention was paid to mitigating the risk of cybersickness. The results indicated that the non-immersive 3D device was more suitable for knowledge retention, whereas the immersive device decreased training time and the number of instructors' assistance required for practical skills acquisition. In addition, the use of the immersive device did not produce significant symptoms of cybersickness. Finally, both devices proved to be easy to use, with a preference for the immersive one.

CCS CONCEPTS

• Human-centered computing → Human Computer Interaction; 300; Applied computing → Education, Training; 300.

KEYWORDS

Virtual reality; Training; Usability; Learning effectiveness.

RÉSUMÉ

L'article rapporte l'évaluation d'un dispositif immersif de réalité virtuelle pour la formation de pilotes. Il est comparé à l'utilisation d'un dispositif de formation 3D non immersif, qui repose sur des écrans plats. Précisément, l'évaluation visait à comparer l'utilisabilité des deux dispositifs en mettant l'accent sur l'efficacité de l'apprentissage. Ainsi, 12 participants ont effectué une tâche d'apprentissage en utilisant les deux dispositifs. Les données recueillies comprenaient la performance d'apprentissage et la satisfaction. En outre, une attention particulière a été accordée à l'atténuation du risque du cybermalaise. Les résultats ont indiqué que le dispositif 3D non immersif était plus adapté à la rétention des connaissances, tandis que le dispositif immersif a diminué le temps de formation et le nombre de demandes d'assistance pour l'acquisition des compétences. De plus, l'utilisation du dispositif immersif n'a pas produit de symptômes significatifs de cybermalaise. Enfin, les deux dispositifs se sont révélés faciles à utiliser, avec une préférence des participants pour le dispositif immersif.

MOTS CLÉS

Réalité virtuelle; Entraînement; Utilisabilité; Efficacité d'apprentissage.

INTRODUCTION AND RATIONALE

In recent years, Virtual Reality (VR) technology has developed considerably. One of its fields of application is training, which has been widely used in various contexts such as patient rehabilitation [1], firing [2], and stress management [3] to name a few. Indeed, VR can effectively simulate different working conditions, and, at the same time, successfully support the learning process (i.e., the process of acquiring knowledge and skills [4]) by doing.

A good example is a study by [5] in which miners used a VR-based training system to enable the acquisition and practice of correct behavior. The results indicated that the 21 participants considered the system useful and felt the positive effects of training even after three months. This encouraged the mining company to introduce VR training into the basic training of junior miners [5]. Similarly, VR has been recognized as useful for training emergency first responders in [6]. The training was conducted with 368 trainees in fully immersive (360 VR) and non-immersive (Desktop VR) environments. The task was a search and rescue operation necessitated by a fire. Overall, the trainees perceived the benefits to far outweigh the costs. In addition, immersive VR was found to be as fit for providing successful training as non-immersive VR. However, immersive VR generated considerably more motion sickness in trainees [6]. This finding highlights the need to consider the potential side effects of VR exposure, particularly cybersickness that causes light to severe discomfort in more than 60% of VR users [7]. This discomfort includes, but is not limited to, headaches, eye strain, nausea, disorientation, and stomach awareness [8].

To summarize, VR training promises to have advantages over traditional training delivered via conventional flat-screen devices. It improves human capabilities and motivation to absorb new content and correct inefficient and wrong work procedures [9]. In addition, it can reduce training time and cost if properly tailored to the trainee's learning style and skills [10].

However, the variability in the quality of VR systems and the risk of potential side effects are sources of limitation that may impact the effectiveness of training [11], i.e., the ability to effectively learn what is targeted by a training program. So, it is important to choose a VR system that is suitable for the training context and trainees' profiles, while meeting safety compliances and maintaining users' well-being. To this end, the suitability of VR-based training should be evaluated through studies that specifically focus on human factors and usability aspects.

Objectives

This article reports on the effects of an immersive VR training device on the pilots' learning effectiveness as defined in [12]. It is compared to a non-immersive 3D training device (used with a laptop) with similar learning goals. The main test objective, then, is to determine whether the VR training device enables trainees to learn effectively during a training task. In addition, special attention is given to monitoring potential side effects, mainly cybersickness.

EXPERIMENTAL DESIGN

Computing Environments

The two training devices include equivalent pedagogical features that enable a trainee to practice a set of Standard Operating Procedures (SOPs) displayed as lessons. Both of them propose gradual levels of guidance to display information. The difference between the devices lies in the way the environment - represented by a 3D Airbus cockpit - is displayed and how to interact with it.

In the **immersive VR device**, the cockpit is visualized with an HP reverb G2 VR head-mounted device (HMD) connected to a VR-capable computer via cable. It provides a resolution of 4320 x 2160 (combined), a 90 Hz refresh rate, and a 114° field of view. Moreover, audio guidance from a synthetic instructor is rendered over the HMD headphone. Finally, the interaction with the cockpit is done through two HP reverb G2 controllers connected to the HMD via Bluetooth.

In the **non-immersive device**, the cockpit is displayed on the flat screen of a laptop with a resolution of 1920 x 1080. The interaction with the cockpit is performed through a mouse and a keyboard.

Task

The training task consisted of learning two pre-flight SOPs containing theoretical knowledge and practical skills, allowing users to improve their cognitive and motor skills.

Thus, both procedures had to be practiced four times by users during an interactive lesson with four levels of guidance, from the most assisted level including audio (e.g. a virtual instructor providing audio instructions), visual (e.g. arrows indicating positions of interest), and textual indications (e.g. a list of written tasks to be performed) to a level of full autonomy without any guidance.

Assessment Tools

The evaluation was based on an objective measure, namely the effectiveness of learning (i.e., acquiring knowledge and skills [4]), and on subjective user satisfaction.

Knowledge acquired during the task was assessed using a theoretical post-questionnaire designed by four Pilot Instructors, while skill acquisition was assessed by a post-check on a procedure training device that physically represents an Airbus cockpit. Regarding users' satisfaction, the usability of both devices was assessed using the System Usability Scale (SUS) [13]. Moreover, a customized post-experience survey assessed users' training preferences and degree of satisfaction.

Finally, to monitor the degree of cybersickness experienced, the Simulator Sickness Questionnaire (SSQ) [14] was administered before and after VR exposure.

Factors

The evaluation followed a within-subject design² with one factor: the **training device** with two conditions labeled **immersive** and **non-immersive**. Moreover, the order of conditions was randomized to control possible learning effects.

USER EVALUATION

Participant sample

Twelve participants took part in the evaluation. All participants had a Private or Commercial Pilot License but they had neither experience with Airbus aircraft nor knowledge of Airbus procedures. All of them freely volunteered for the evaluation, without any financial compensation.

There were three women and nine men with ages ranging from 24 to 50 years (mean age = 38.8, SD = 7.1). Five participants reported having no experience with VR technology, while the others reported having used VR at least once. Two out of nine participants used their glasses during the VR session. Finally, it is worth mentioning that although interpupillary distance (IPD) was measured (using a pupillometer), it did not constrain the selection of participants. Thus, three participants had an IPD lower than the minimum value of the HMD (mean IPD = 61.2, SD= 2.99).

² Note that the effectiveness of learning was assessed during the 1st condition only (see "Test procedure" subsection).

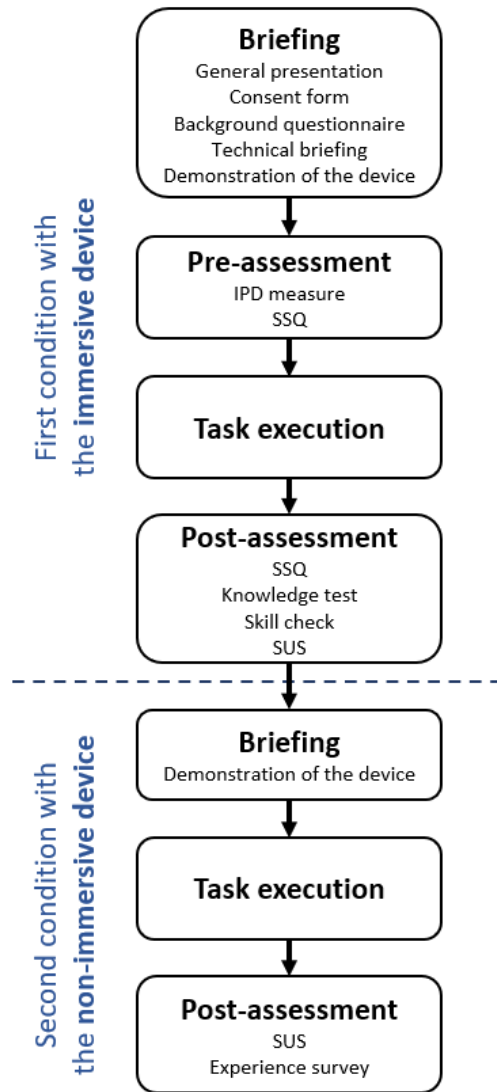


Figure 1: Protocol timeline with the immersive device in the first condition, and the non-immersive device in the second condition.

Test procedure

Upon their arrival, participants were welcomed by the evaluation team of three experimenters. They signed an informed consent form and filled out a background information document. The test was divided into four phases as follows (see Figure 1):

Phase 1 – Briefing [Average duration = 57 min]. It included a presentation of the training devices, the VR equipment cleaning protocol, the potential risks associated with VR use, the learning task to perform, and an aircraft systems and cockpit overview. Then, an experimenter demonstrated how to use the training device and participants took a 10-minute break.

Phase 2 - Pre-assessment [Average duration = 2 min]. In the immersive condition only, the participants filled out the SSQ and their IPD was measured.

Phase 3 – Task execution [Average duration = 27 min]. Participants undertook the discovery of the corresponding device for about 5 minutes. Once comfortable, they performed the task. Note that the participants were not allowed to remain immersed for more than 23 minutes in a row to mitigate cybersickness. The duration was tracked by an experimenter.

Phase 4 - Post-assessment [Average duration = 19 min]. Once the task was completed, participants filled out the SSQ (only in the immersive condition). Then, they passed the knowledge post-test and the skills post-check on the procedure training device. This was done only in the first device condition, whether it was the immersive or non-immersive condition. Furthermore, the SUS questionnaire was administered. Then, participants took a 10-minute break before performing the task using the second device.

Finally, after completing the second condition, they filled out the post-experience survey to provide their preferences. The total duration of the evaluation was about 3 hours.

Data collection

In addition to collecting learning and satisfaction metrics using dedicated tools (see “Assessment Tool” section), different performance measures were tracked during the task as follows:

- **Task completion time**, records in seconds.
- **Number of requests for assistance** refers to instances where an instructor had to assist a participant. All the participants were assisted during the test.
- **Number of attempts** refers to instances where participants had to perform a step more than once. For example, not being able to point to a small area on the first try.
- **Number of mistakes** refers to instances where participants made an error in the task. For example, taking the wrong action.

Evaluation of Virtual Reality Training

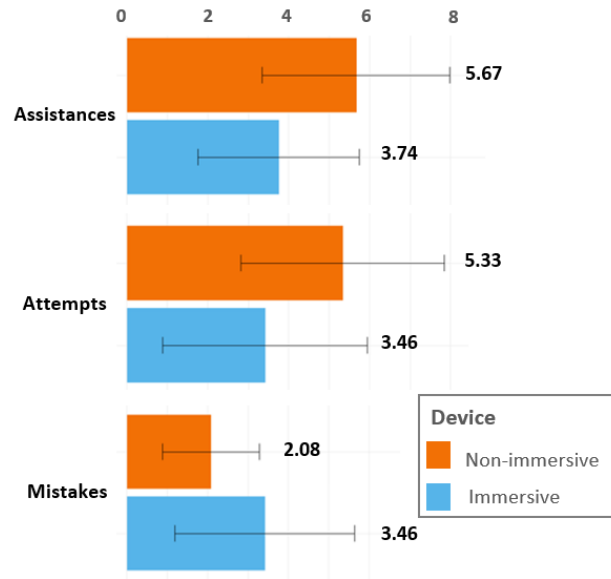


Figure 2: Mean performance scores for immersive and non-immersive conditions.

Condition	Non-immersive	Immersive
Knowledge score	83%	47%
Skill score	89%	79%
Completion time	28min 25sec	16min 28sec
Knowledge/time	2.96	2.94
Skill/time	3.18	4.94

Table 1: Knowledge score, skill score, and completion times mean for immersive and non-immersive conditions.

RESULTS

Cybersickness measure

The average SSQ score was 6.54 immediately post-exposure, which is associated with low physiological impacts [15]. In addition, no statistically significant effect of VR exposure on SSQ score was found (Wilcoxon signed-rank test: $p = .28$, Cohen's $d = .39$). Furthermore, only one participant reported dry eyes after-effect, which disappeared after one night of sleep.

Performance measure

Figure 2 displays the mean scores of the number of assistance, attempts, and mistakes for both conditions. The results indicated that participants were less assisted and made fewer attempts in the immersive condition. However, they made more mistakes. These mistakes were mainly sequencing errors (ex. doing an action before another) or forgetting errors (ex. not doing an action).

Learning measures

Table 1 shows the averages of knowledge scores, skill scores, task completion time, and the relationship between these values. Thus, knowledge and skill scores were lower on the immersive device (with a reduced gap in skill acquisition). However, these scores were achieved in less time. In particular, the relationship between learning rate and completion time showed a greater efficiency of the immersive device concerning skill acquisition.

Satisfaction measures

The SUS scores indicate that the immersive device was rated as better (80.6pts) than the non-immersive device (73.75pts) in terms of usability. Yet, no significant difference was found (Wilcoxon signed-rank test, $p = .06$, Cohen's $d = .63$) and both devices provide a "good" perceived usability according to [16].

In contrast, a statistically significant effect of session order was found on SUS scores for the immersive condition (Wilcoxon signed-rank test, $p < .05$, Cohen's $d = 2.97$), see Figure 3, pushing the immersive condition, in the second session, to an "excellent" perceived usability according to [16].

In addition, responses to the post-survey provided information about participants' preferences:

- 100% preferred to use the immersive device for skill learning.
- 83% preferred to use the immersive device for global learning.
- 25% preferred to use the immersive device for knowledge learning.

Moreover, half of the participants proposed to use non-immersive and immersive devices in a complementary manner, the former for knowledge retention, and the latter for skills practice.

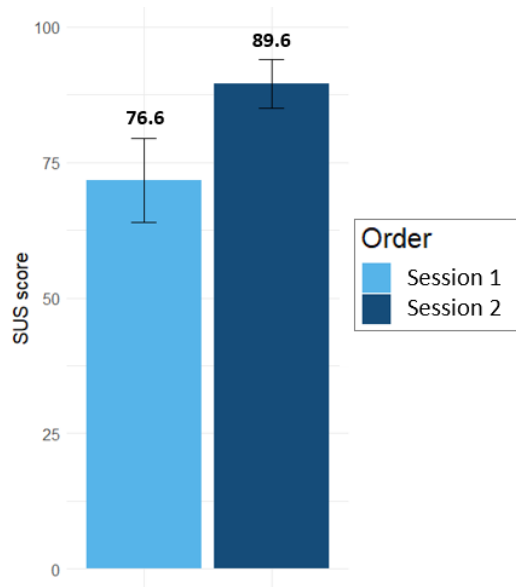


Figure 3: SUS score mean in immersive condition between first and second session order. The score significantly increases when the VR device was used after the non-immersive device.

DISCUSSION

This user evaluation focused on assessing whether the immersive VR training device enables trainees to learn effectively while performing a training task (in this case, two pre-flight procedures).

On one side, the results indicated that the immersive device was less effective for knowledge retention than the non-immersive device. These results are consistent with previous research reporting that VR may not be superior to conventional flat-screen media for acquiring certain intellectual knowledge [17, 18].

On the other side, results between the non-immersive and the immersive devices were closer regarding skill performance. Yet, participants reached these skill scores more than 1.75 times faster using VR and with less assistance required from the instructors, which is a significant improvement in a training program. This can be explained by the naturalness of the immersive virtual cockpit and the affordance of its interactions. Thus, applying the gestures associated with each action was easier than using a mouse on a flat-screen computer. The fact that the participants needed fewer attempts to perform each action also indicates ease of interaction. In addition, the different levels of guidance in the immersive device were sufficient to allow participants to gain autonomy faster. As a result, fewer interventions from instructors were needed to help them complete their learning.

Furthermore, this study showed how non-immersive and immersive devices could be used in a complementary manner to optimize different training acquisitions, the former for knowledge retention, such as knowing which engine number to start first in an aircraft, and the latter for skill practice, such as being able to interact correctly with the appropriate controls.

Finally, regarding the potential cybersickness risk, the VR device with its current pedagogical content and controlled exposure time successfully maintained the users' well-being.

CONCLUSION AND PERSPECTIVES

The user evaluation presented in this article showed good usability of the VR immersive device for pilot training. In particular, the VR device demonstrated its effectiveness to initiate skill acquisition of a procedure application in a reduced time. These results open up promising perspectives for the use of such a device as support for skill practice that would outperform existing means.

Thus, further studies should enable the evaluation of the learning efficiency of the VR device (performance versus cost) in an operational pilot training program. In addition, the exploration of session duration extension (i.e. longer than 23 minutes) will allow defining enhanced recommendations to maximize user well-being and training effectiveness.

ACKNOWLEDGMENTS

Thanks to all the participants for the quality of their participation, to the Airbus stakeholders (Instructors, Flight Operations department, VR project team) for the intensity of the exchanges, and to the French Civil Aviation Authority (DGAC) for supporting this study.

REFERENCES

- [1] Rossol, N., Cheng, I., Bischof, W. F., & Basu, A. (2011, December). A framework for adaptive training and games in virtual reality rehabilitation environments. In *Proceedings of the 10th international conference on virtual reality continuum and its applications in industry* (pp. 343-346).
- [2] Bhagat, K. K., Liou, W. K., & Chang, C. Y. (2016). A cost-effective interactive 3D virtual reality system applied to military live firing training. *Virtual Reality*, 20(2), 127-140.
- [3] Pallavicini, F., Argenton, L., Toniazzi, N., Aceti, L., & Mantovani, F. (2016). Virtual reality applications for stress management training in the military. *Aerospace medicine and human performance*, 87(12), 1021-1030.
- [4] Antonacopoulou, E. P. (2001). The paradoxical nature of the relationship between training and learning. *Journal of Management Studies*, 38(3), 327-350.
- [5] Grabowski, A., & Jankowski, J. (2015). Virtual reality-based pilot training for underground coal miners. *Safety Science*, 72, 310-314.
- [6] Pedram, S., Ogie, R., Palmisano, S., Farrelly, M., & Perez, P. (2021) Cost-benefit analysis of virtual reality-based training for emergency rescue workers: a socio-technical systems approach. *Virtual Reality*, 25(4), 1071-1086.
- [7] B. Lawson. Motion Sickness Symptomatology and Origins, pp. 531– 600. 09 2014. DOI: 10.1201/b17360-29.
- [8] J. J. LaViola, Jr. A discussion of cybersickness in virtual environments. *SIGCHI Bull.*, 32(1):47–56, Jan. 2000. DOI: 10.1145/333329.333344
- [9] Podgórski, D. (2010). The use of tacit knowledge in occupational safety and health management systems. *International Journal of Occupational Safety and Ergonomics*, 16(3), 283-310.
- [10] Iván Aguilar Reyes, C., Wozniak, D., Ham, A., & Zahabi, M. (2022, September). An Adaptive Virtual Reality-Based Training System for Pilots. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 66, No. 1, pp. 1962-1966). Sage CA: Los Angeles, CA: SAGE Publications.
- [11] Kaplan, A. D., Cruik, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of VR, AR, and mixed reality as training enhancement methods: A meta-analysis. *Human factors*, 63(4), 706-726.
- [12] International Organization for Standardization. ISO 9241-11:1998, Ergonomic requirements for office work with visual display terminals (VDTs), Part 11: Guidance on usability. Geneva, Switzerland: Author. 1998.
- [13] Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- [14] Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3), 203-220
- [15] Stanney, K., Lawson, B. D., Rokers, B., Dennison, M., Fidopiastis, C., Stoffregen, T., ... & Fulvio, J. M. (2020). Identifying causes of and solutions for cybersickness in immersive technology: reformulation of a research and development agenda. *International Journal of Human-Computer Interaction*, 36(19), 1783-1803.
- [16] Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114-123.

- [17]** Babu, S. K., Krishna, S., Unnikrishnan, R., & Bhavani, R. R. (2018, July). Virtual reality learning environments for vocational education: A comparison study with conventional instructional media on knowledge retention. In *2018 IEEE 18th international conference on advanced learning technologies (ICALT)* (pp. 385-389). IEEE.
- [18]** Hall, C. R., Stiles, R. J., & Horwitz, C. D. (1998, March). Virtual reality for training: Evaluating knowledge retention. In *proceedings. IEEE 1998 virtual reality annual international symposium (Cat. No. 98CB36180)* (pp. 184-189). IEEE.