HAL
open science

Eyestrain impacts on learning job interview with a serious game in virtual reality

Alexis D Souchet, Stéphanie Philippe, Dimitri Zobel, Floriane Ober, Aurélien Lévěque, Laure Leroy

To cite this version:

Alexis D Souchet, Stéphanie Philippe, Dimitri Zobel, Floriane Ober, Aurélien Lévěque, et al.. Eyestrain impacts on learning job interview with a serious game in virtual reality. VRST '18: 24th ACM Symposium on Virtual Reality Software and Technology, Nov 2018, Tokyo Japan, France. pp.1-12, 10.1145/3281505.3281509 . hal-03532393

HAL Id: hal-03532393

https://hal.science/hal-03532393v1

Submitted on 18 Jan 2022

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.

Eyestrain Impacts on Learning Job Interview with a Serious Game in Virtual Reality

A randomized double-blinded study†

Alexis D. Souchet
Paragraphe Laboratory
Paris 8 University & Manzalab
Paris, France
asouchet@manzalab.com

Stéphanie Philippe
R&D Department
Manzalab
Paris, France
sphilippe@manzalab.com

Dimitri Zobel
University Paris-Est Créteil
Créteil, France

Floriane Ober
Manzavision
Aix-en-Provence, France

Aurélien Lévêque
Manzavision
Aix-en-Provence, France

Laure Leroy
Paragraphe Laboratory
Paris 8 University
Saint-Denis, France
laureleroyrv@gmail.com

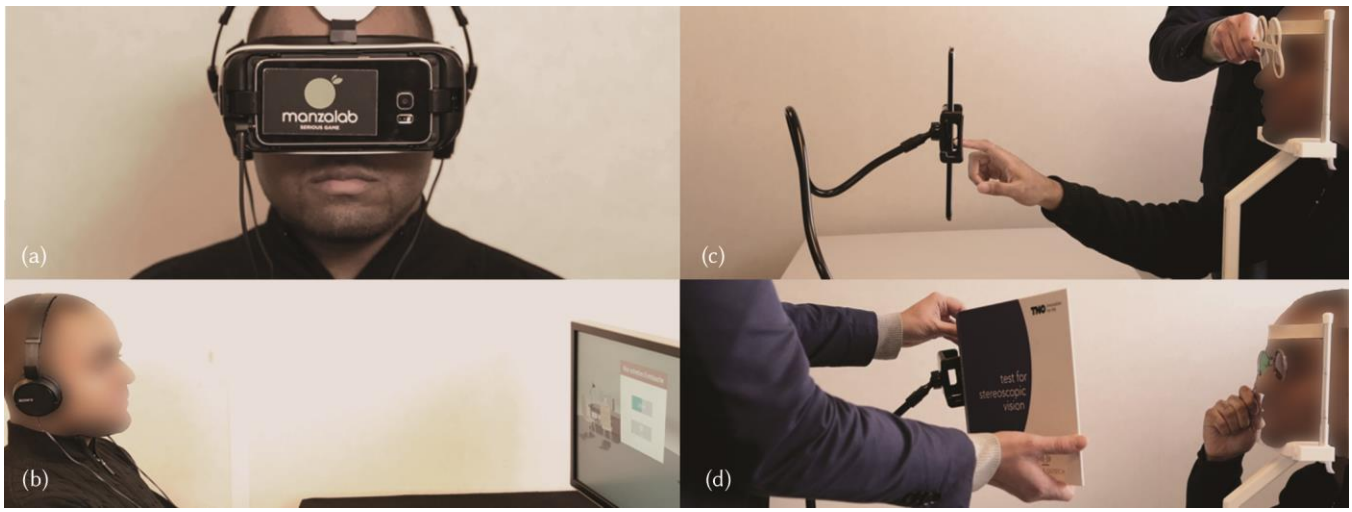

Figure 1: (a) Subject wearing the HMD for condition A and B, (b) Subject in-front of computer screen with mouse for condition C, (c) Performing the Flipper lens test (ease of accommodation), (d) Performing the TNO test (stereo acuity)

ABSTRACT

Purpose: This study explores eyestrain and its possible impacts on learning performances and quality of experience using different apparatuses and imaging. Materials and Methods: 69 participants played a serious game simulating a job interview with a Samsung Gear VR Head Mounted Display (HMD) or a computer screen. The study was conducted according to a double-blinded protocol. Participants were randomly assigned to 3 groups: PC, HMD biocular and HMD stereoscopy (S3D). Participants played the game twice, allowing between group analyses. Eyestrain was assessed pre- and post-exposure on a chin-head rest with optometric measures. Learning traces were obtained in-game by registering response time and scores. Quality of experience was measured with questionnaires assessing Presence, Flow and Visual Comfort. Results: eyestrain was significantly higher with HMDs than PC based on Punctum Proximum of accommodation and visual acuity variables and tends to be higher with S3D. Learning was more efficient in HMDs conditions based on time for answering but the group with stereoscopy performed lower than the binocular imaging one. Quality of Experience was better based on visual discomfort with the PC condition than with HMDs. Conclusion: learning expected answers from a job interview is more efficient while using HMDs than a computer screen. However, eyestrain tends to be higher while using HMDs and S3D. The quality of experience was also negatively impacted with HMDs compared to computer screen. Not using S3D or lowering its impact should be explored to provide comfortable learning experience.[1]

[1]Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

VRST '18,November 29-December 1, 2018, Tokyo, Japan© 2018 Association for Computing Machinery. ACM ISBN 978-1-4503-6086-9/18/11… $15.00
https://doi.org/10.1145/3281505.3281509

CCS CONCEPTS

• Human-centered computing →Interaction paradigms; Virtual reality

KEYWORDS

Virtual Reality, Eyestrain, Serious Game, Stereoscopy, Head Mounted Display, Learning

ACM Reference format:

Alexis D. Souchet, Stéphanie Philippe, Dimitri Zobel, Floriane Ober, Aurélien Lévêque and Laure Leroy. 2018. Eyestrain Impacts on Learning Job Interview with a Serious Game in Virtual Reality: a Randomized Double-blinded Study. In Proceedings of ACM VRST conference (VRST'18). ACM, Tokyo, Japan. https://doi.org/10.1145/3281505.3281509

1 INTRODUCTION

The use of current Head Mounted Displays (HMDs), regardless of offering a serious game may lead to risks for users. Such devices raise particular problems associated with human factors and more particularly with the human visual system [19, 33]. HMDs can be used to present stereoscopic contents (S3D). Stereoscopy attempts to reproduce depth cues perceived by the human visual system under "natural" conditions [2]. However, displaying S3D imaging can lead to eyestrain for users [47]. The combination of serious game (SG) with virtual reality (VR) and S3D raises an issue about the possible side effects on learning performance and on the human visual system in the context of learning. Previous works usually addressed task efficiency for learning purposes without considering possible effects of eyestrain on general learning curves. Therefore, linking possible eyestrain effects on learning should be investigated. However, data about risks of eyestrain mostly concentrate on late nineties' technologies and lacks data based on modern HMDs. Such HMDs are increasingly sold to the general public, raising the needs to the compile of more recent data.

2 RELATED WORK

2.1. Learning efficiency and stereoscopy

Previous studies have attempted to identify the effectiveness of learning when S3D is displayed. For example, Keebler indicates that S3D conditions would negatively impact retention [18]. Indeed, the S3D would be a distraction in some cases. This tendency is found in other cases with the S3D less efficient than 2D [31]. A possible reason for such results evoked by Mukai *et al.* would be the eyestrain reported by subjects in S3D.

On the contrary, according to the study by Kulshreshth *et al.* S3D (television) would improve scores in video games and offer a better learning curve for succeeding in video game tasks [21]. In line with this, Loup-Escande *et al.* indicate that when performing tasks, the S3D leads to better efficiency for visual search [23]. However, their evaluation of learning after exposure shows no significant difference between S3D and monoscopic cues. In addition, the subjective immersion and flow experience is better in the S3D condition. Learning surgical technics have also been investigated and often shown to be better in S3D than with monoscopic cues only [10]. A limitation has however been reported in the case of novice learners: S3D would add a cognitive load driving the 2D condition to more efficiency [28].

In summary, the benefits of S3D compared to monoscopic cues seem to depend on the type of tasks to be performed [26], the level of expertise of the users and the characteristics of the proposed virtual environment. Given these conditions, it seems difficult to extend the results of the studies presented above to all learning configurations.

2.2. Learning efficiency combination SG-VR-S3D

Some tasks, including learning could benefit from the association SG-VR-S3D [3]. We include the simulations in this review. Stone *et al.* conduct a review of ten virtual environments and report an efficient learning, associated with an attractive training cost and a short training time [43]. For other kind of learning, passing a forklift driving license test for example, coupling SG-VR via HMD could have positive results on the quality of experience of learners compared to computer screen (PC) [15]. This result is consistent with Webster's study including 140 trainees learning prevention and control of corrosion of materials [49]. His results indicate better learning with VR than with the reference training method. It is however important to note that, pointed by Webster, the design of instructions between the two conditions was significantly different to the advantage of the virtual environment. According to the study by Schmoll *et al.*, short-term retention is better in VR than with a computer screen and mouse. This effect was however not maintained in time as the authors note a lower retention over the long term in VR condition [39].

In summary, it seems that the association SG-VR-S3D presents disparate results. The differences that are reported could be due to the fact that learning instructions are different from one device to another, not to the device itself.

2.3. Eyestrain in VR with stereoscopy

Previous studies point out the conflict between vergence and accommodation as a determining factor in triggering an eyestrain when S3D content is displayed [14, 48]. Under natural conditions, these two proprioceptive cues are coactive in their dynamic treatment [40]. However, the attempt to imitate stereopsy via stereoscopic imaging makes the two processes disassociated [8, 24]. Within HMDs the accommodation is fixed because the human visual system is focus on the plane of the screens' projection while the vergence is focused on the object of interest. The HMD screens are very close to the visual system and it is thanks to the lenses that the

focal plane is offset. Moreover, it seems that the cognitive demand for the tasks to be performed would have an impact on accommodation [16]. However, in a virtual environment such as a SG-VR, cognitive treatments are numerous.

Kulshreshth *et al.* previously reported that eyestrain did not have a negative impact on subject performances [21]. However, this result is based solely on the subjects' statements. Eyestrain has been observed after the use of HMDs and S3D in many cases [29, 30, 36, 46]. This phenomenon still seems to be topical with the latest HMD models [12]. However, through our literature review, eyestrain is given little consideration when evaluating the effectiveness of the SG-VR-S3D combination. Learning tasks are not preferred when evaluating the eyestrain and its impacts. It should be noted that learning tasks are indeed proposed in the literature dealing with simulations. However, these tasks are rarely in ecological conditions and involve little the characteristics of the devices currently available.

2.4. Quality of experience with virtual environments and Learning

One of the markers of learning effectiveness in training is the degree of transfer [5]. According to Alexander *et al.* four factors influence it: fidelity (of the virtual environment), immersion, presence and user acceptance [1]. According to Slater and Wilbur, presence corresponds to an individual psychological consciousness state having the feeling of "being there" in the virtual environment [42]. Thus, immersion influences the degree of presence and presence is an indicator of immersion quality [37] proposed by the apparatus [7, 27]. Flow, introduced by Csikszentmihalyi, corresponds to an optimal state of intrinsic motivation mobilizing all the resources of a person towards the action undertaken, finding oneself totally absorbed and to the maximum of her/his capacities [6]. Flow can be used as a proof of engagement according to previous work.

According to the study by Bowman *et al.*, the greater the degree of immersion, the better the subjects' ability to remember information [4]. Hamari *et al.* based on two experiments conclude that the higher the engagement, the more effective the learning, but the reverse is true for immersion as higher immersion correlates with less efficient learning [13]. Limperos *et al.*, indicate that non-expert players are more likely to report a greater presence and immersion in S3D conditions compared to experts [22]. In one of Freeman *et al.*'s study, S3D leads to a higher degree of presence than monoscopic cues [11]. Takatalo *et al.* find, on the contrary, no significantly better results for the flow and involvement of subjects playing in S3D [44]. According to Pavlas *et al.*, a good flow would positively impact learning via SG, which would be a predictor of learning reported by students [34]. Thalmann *et al.*, indicate that spatial presence and social presence are better in CAVE and Oculus rift DK2 [45]. Schrader and Bastiaens indicate that the presence was higher with the SG than with a hypertext interface and according to the authors, the presence could have a link with attention [41].

2.5. Considering eyestrain while learning with HMDs

The consideration of HMDs used by the general public, as a device for displaying S3D, is often absent from reviews [17]. Thus, studies integrating HMDs on S3D and learning issues are quite rare. This can be explained by their very recent reappearance on the general market. Yet, the quality of experience and efficiency of task performance would be different with HMD compared to a computer and this should be further investigated. The study by Santos *et al.* suggests a better efficiency in performing navigation tasks (via video game) in computer than in HMD [38]. The authors point out the difficulties caused by the HMD cables and the degree of expertise of the subjects with video games as possible explanations for their results. It should be noted that although sometimes mentioned in the studies mentioned above, eyestrain remains little considered and measured. Previous studies have indicated a difference between computer screen and HMD in terms of eyestrain [29, 30, 36, 46]. Thus, the interest of presenting images in stereoscopy when learning via SG in HMD-VR remains an issue requiring more investigation because: 1) users could present eyestrain, 2) attention could be less concentrated on educational content because of the presence of S3D, 3) eyestrain could influence cognitive fatigue thus lowering attention faculties particularly in a context of high cognitive load such as SG in VR. However, S3D may be necessary for optimal task performance in the virtual environment or learning involving the centrality of visuo-spatial cognitive processing [32, 35].

The state of the art regarding the association of stereoscopy, serious game and virtual reality has led us to identify a lack of qualification of the limits of use, as regards eyestrain as well as learning performances. Moreover, the influence of eyestrain, psychomotor limitations and perception specific to these devices on learning abilities need more documentation. Our study proposes an empirical contribution on these aspects.

3 MATERIAL AND METHODS

This study proposes to expose subjects to a serious game in virtual reality by varying the device (PC or HMD) and the imaging proposed: biocular (corresponding to the same image displayed on both), or stereoscopy (corresponding to two different images displayed on both screens). The subjects are randomly divided into 3 groups:

- Group A with a Samsung Gear VR with S6 displaying biocular imaging (HMD-Biocular),
- Group B with a Samsung Gear VR with S6 displaying binocular imaging (stereoscopy) based on Oculus recommendations (HMD-S3D),
- Group C (control group) with a computer screen.

The steps of this protocol are performed in accordance with the methodology of double-blinded investigation and between-groups design.

3.1. Ethical aspects and recruitment

The subjects were volunteers. This research protocol has been reviewed and approved by the CER – Paris-Descartes (independent Research Ethics Committee), registered as: (2017-49) 2018-18. This study has been declared to the CNIL (National Commission on Informatics and Liberty) relating to data collection and protection registered which validated conformity with the "MR1 - Health research with consent" methodology, registered as: 2120463 v 0. An information letter indicated the aims of the study to the volunteers and a consent form was signed by the subjects before starting the experiment. Recruitment was done online via social networks and an email mailing list (RISC by the CNRS). A € 20 voucher compensation for each subject was proposed.

3.2. Serious Game in VR used

This serious game aims at training people to job interview (Fig. 2-3). The learning instructions have been built based on the confidential standards of Human Resources of Pôle Emploi (French governmental agency: employment centre). The original game offers several scenarios and is called "Mon entretien d'embauche" (my job interview). For reproducibility purpose of the study, we used a unique scenario based on a salesperson position.

The subject chooses an avatar between a male and a female. The subject then takes notice of the job offer and the resume of the avatar. Then, the scenario consists of: 1) Discussion with a friend before the interview, woman or man according to the choice of avatar, 2) Reading the company's website, 3) Interview with the recruiter, 4) Review of the interview with a coach. The interactions are in the form of "Point and click" with dialogs. Dialog voices have been registered with professional actors. The player can turn the camera at 360°. Subject has a first-person viewpoint. The player does not see his/her avatar.

The goal of the game is to choose the most relevant answer, usually between four choices of dialogue. Each choice is weighted according to the relevance to the question asked, the context, and the standard codes of the job interview. A score is displayed in real time and changes according to the player's choice. The answer is selected by hovering one of the dialogue boxes: with a HMD by placing the pointer while moving the head, by mouseover with a PC. A loading time validates the selection. The time to answer is limited: a loading bar symbolizes the time remaining. In addition, a direct feedback is given when a response is chosen: red = unsuitable answer, orange = average answer, green = expected answer. In the third act, the coach comments the strengths and weaknesses of the player during the interview. The final score is displayed and a percentage of achievement in the three skills is listed: marketing = ability to "sell oneself," communication = ability to communicate according to expected codes, conduct = ability to lead the interview. This allows the learner to identify possible improvements.

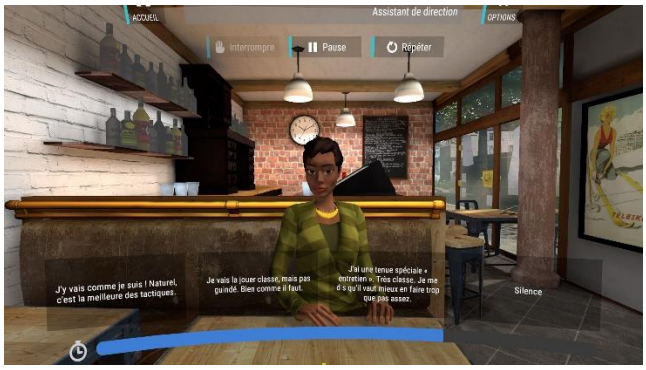

Figure 2: screenshot from the SG "My job interview" in front of the friend (act 1) © Manzalab

3.3. Apparatus

3.3.1. Samsung Gear VR™ characteristics (group A and B). Model: SM-R321; This HMD requires a smartphone inserted. Its technical characteristics are as follows: Field Of View: 90°; tracking: 4 DoF (up, down, left, right) with Accelerometer and Gyroscope in the Galaxy S6; weight: 345 grams.

3.3.2. Samsung© Galaxy 6™ characteristics. OS: Android 8.1; Processor: Samsung Exynos 7 Octa @ 2,1 GHz; graphic card: ARM Mali T760; memory: 3 GB (RAM); display: 5.1" Quad HD Super AMOLED. 2560 x 1440 px (577 ppi); luminance: 432 cd/m^2.

3.3.3. PC and display Dell© characteristics (group C). OS: Windows 10 Family version 1709; Processor: i7-4790 @ 3.60 GHz; memory: 8 GB (RAM); graphic card: Nvidia GeForce GTX770 2GB; Display model: Dell© 2407WFP-HC Flat Panel Monitor; Display: 24" Active matrix-TFT LCD. 1920 x 1200 px; Refresh Rate 81 kHz (horizontal), 76 Hz (Vertical); luminance: 400 cd/m^2.

3.3.4. Headphones Sony© (all groups). Model: MDR-ZX110B. Frequency: 12 to 22 kHz; diaphragm: 30 millimeters; weight: 120 grams.

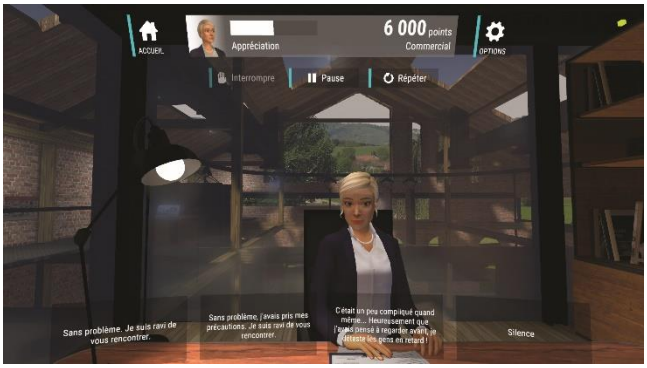

Figure 3: screenshot from the SG " My job interview " in front of the recruiter (act 2) © Manzalab

3.4. Participants

69 subjects took part in this study. The sample consisted of 41 women and 28 men aged 18 to 39, M = 25.01 ± 0.21 (SD = 5.30) years. Subjects were excluded if: over the age of 40 (presbyopia can influence optometric tests), declaring a pathology greatly influencing the state of the visual system and perceptions. The subjects were mainly students (from various fields: psychology, digital, human resources, business, marketing and arts) as well as researchers (in psychology for the most part). 37 subjects had no correction and 32 had either glasses or correction lenses. Of these, 21 had myopia, 2 astigmatism, 7 combined myopia and astigmatism, and 2 combined hyperopia and astigmatism.

3.5. Procedure and measures

The subject completes a profile questionnaire.

Eyestrain evaluation is based on optometric measurements performed before (pre-) and after (post-) exposure at 40 cm from the subject with a head-chin rest (Punctum proximum of accommodation and stereoscopic acuity) and with the Smartoptometry© application on a Samsung Galaxy tablet (ease of accommodation and visual acuity):

- Punctum proximum of accommodation, measured with the Donder's Push-up Test, then clustered at a speed of 5 to 7.5 cm per second,
- Stereoscopic acuity, measured with the TNO test,
- Ease of accommodation, measured via the Flipper lens test (+ 2.00 / -2.00) for 1 minute with the Smartoptometry© feature displaying words on a tablet that participants had to swipe when seen in-focus,
- Visual acuity, measured by detecting the E of Raskin (random) orientation of smaller and smaller dimensions.

Learning curves are based on learning performance measures obtained during the game: scores and response time as markers of memorization, low values of score correspond to low learning performance and low response time correspond to high learning performance.

The quality of experience is measured via questionnaires with Likert scales graduated from 1 to 5 at the end of the experiment (translated into French by the first author):

- Visual discomfort: measured via the questionnaire by Zeri and Livi consisting in 11 items [50].
- Presence: measured via the Multimodal Presence Scale by Makransky *et al.* with the spatial and social presence items being 10 items because the self-presence part is not relevant given the terms of our SG [25],
- Flow: measured via the Flow Short Scale by Rheinberg, Vollmeyer and Engeser consisting in 9 items [9].

The subject balance was assessed with the Tinetti-Poma test at the end of the experiment.

The procedure is as follows: 1) Subjects signed the consent form after receiving the information about the study. 2) Optometric measurements (pre-) were conducted by the first author. 3) Subjects went to another room with the third author, invisible to the first author, and received explanation of the game's purposes and interactions, the third author did not know the tested conditions and was guiding subjects and helping to get equipped by following a rigorously similar script for each condition not making variations in instructions. The third author was trained to monitor subjects' wellbeing during exposure. 4) Subjects were exposed for **30 minutes,** a reference time in the literature [20]. The **subjects sit on a chair during exposure** and **play two games** (G1 and G2) of the SG each lasting 15 minutes. **Group C sat at 1 meter from the screen.** Each group wore the same headphones for sound. 5) Subjects came back a few seconds after their exposure and optometric measurements (post-) were performed. 6) Subjects completed Quality of Experience questionnaires. 7) The Tinetti-Poma test was performed to ensure subjects were able to walk out of the building with perfect balance reflexes.

3.6. Analysis and statistics

The random assignation resulted: group A, 24 subjects; group B, 22 subjects; group C, 23 subjects. Tests are carried out on one hand in each group and on the other hand between groups. The significance level is tested with a **confidence of $\alpha = 0.05$**. Jamovi version 0.9.1.9 (2018) was used for statistical tests.

3.7. Hypotheses

H1: learning curves of expected answers in a job interview in a SG is less efficient when the learners present eyestrain.

H2: eyestrain deteriorates the quality of experience in a SG.

H3: S3D drives to more eyestrain than biocular imaging.

H4: learning curves of expected answers in a job interview in a SG is higher with a VR-HMD than with a computer screen.

4 RESULTS

4.1. Eyestrain

4.1.1. Punctum Proximum of Accommodation (PPA). The higher the Donder's Push-up Test score, the lower is the PPA. The Shapiro-Wilk test was performed to test the distribution of each group. Only group B pre- ($p = 0.215$) and post-exposure ($p = 0.667$) data follow a normal distribution, not for the two other groups. In order to test the between groups' data, we subtracted the scores pre- and post-exposure, the results were tested: group B ($p = 0.203$) data follow a normal distribution but not the other groups.

4.1.1.1.Difference pre-/post-exposure within each group. We observe that all 3 groups have increase test scores (Fig. 4). The data are paired. A Wilcoxon Sign Rank Test indicated: for group A, that the

median pre-exposure ranks, MD = -14.50, was statistically different from the median post-exposure ranks Z = -4.197, **p = < 0.001**. It can thus be seen that the PPA of group A fell significantly between pre- and post-exposure; for group C, the median pre-exposure ranks, MD = -14, was statistically different from the median post-exposure rank Z = -3.808, **p = < 0.001**. It can thus be seen that the PPA of group C fell significantly between pre- and post-exposure; for group B a t-test indicated a significant difference in the scores for pre- (M = 68, SD = 3.46) and post-exposure (M = 82.8, SD = 4.05), t (21) = -7.16, **p = < 0.001**. It can thus be seen that the PPA of groups A, B and C significantly fell between pre- and post-exposure, revealing a negative impact of exposure in all 3 conditions.

4.1.1.2.Difference between groups. The data are independent. The Kruskal-Wallis Test indicated that there was a significant difference in PPA pre- and post-exposure between groups (H(19.82) = 2, **p = < 0.001**), with a mean rank of 29.083 for group A, 25.227 for group B and 11.435 for group C (*Fig. 4*). Dunn's Post-Hoc tests without correction were conducted to obtain details of significance for each group compared. The difference between group A and B was not significant (MD = 0.932, SE = 2,880) Z = -0.156, p = 0.438. The difference between group A and C was significant (MD = 10.967, SE = 2,847) Z = 3.818, **p = < 0.001**. The difference between group B and C was significant (MD = 10.036, SE = 2,910) Z = 3.890, **p = < 0.001**.

The post-hoc comparison of the groups then shows a significant difference between the test groups (A and B) and the control group (C), supporting the idea of a negative impact of HMD exposure on the PPA when compared to a computer.

Taken together, the analysis of the PPA measurements in each group before and after exposure therefore supports H3.

4.1.2. Ease of Accommodation (EoA). The Shapiro-Wilk test was performed to test the distribution of each group. Only group C pre- (p = 0.231) and post-exposure (p = 0.168) data follow a normal distribution, not for the two other groups. In order to test the between groups' data, we subtracted the scores pre- and post-exposure, the results were tested: none followed a normal distribution. The lower the Flipper lens test's score, the lower the EoA.

4.1.2.1.Difference pre-/post-exposure within each group. EoA of group A fell slightly by -3.19%, as well as group B by -9.59% and group C by -1.86%, on average between pre- and post-exposure.

The data are paired. A two-tailed Wilcoxon Sign Rank Test indicated for group A, that the median pre-exposure ranks, MD = 1.50, was not statistically different from the median post-exposure ranks W = 58.5, p = 0.084. A Wilcoxon Sign Rank Test indicated: for group B that the median pre-exposure ranks, MD = 2.50, was statistically different from the median post-exposure ranks Z = 2.374, **p = 0.018**; for group C, a t-test indicated no significant difference in the scores for pre- (M = 30, SD = 10.4) and post-exposure (M = 32, SD = 10.5), t (22) = -1.13, p = 0.273. It can be seen that the EoA of groups A, B and C slightly fell between pre- and post-exposure, although the difference is statistically significant only for group B.

4.1.2.2.Difference between groups. The data are independent. The Kruskal-Wallis Test indicated that there was not a significant difference in EoA pre- and post-exposure between groups (H(1.87) = 2, p = 0,392), with a mean rank of 10.25 for group A, 14.773 for group B and 7.130 for group C.

As a conclusion regarding eyestrain, the medians tendency indicates that condition B was the most tiring followed by condition A and finally condition C, although the results are statistically significant only for group B. Therefore, H3 can be supported via this optometric variable but only for the pre-/post-exposure comparison of group B.

Therefore, H3 can be supported via this optometric variable but only for the pre-/post-exposure comparison of group B and based on of medians tendency. However, no statistically significant results corroborate this tendency observed when the difference between groups is tested.

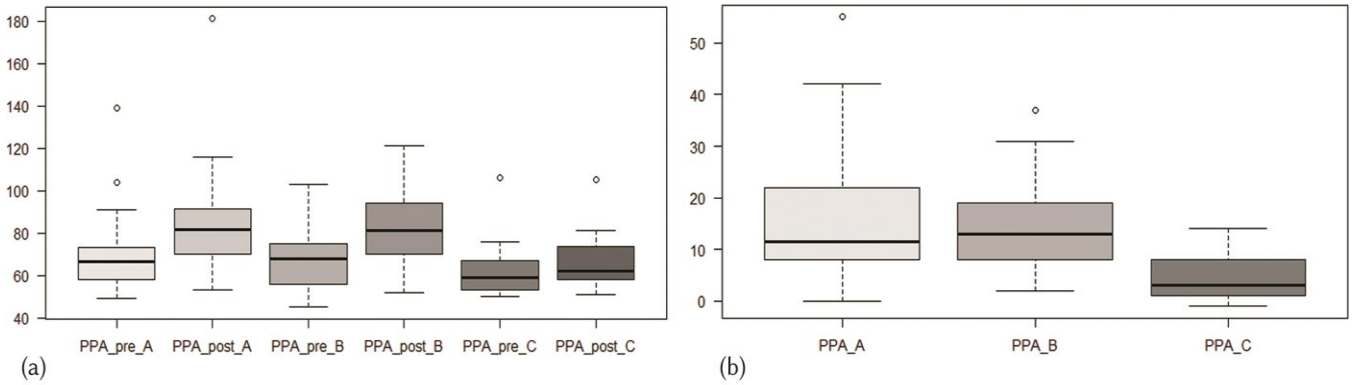

Figure 4: Punctum Proximum of Accommodation (PPA), (a) PPA pre- and post-exposure by group, lighter colors are for pre-exposure and darker colors are for post-exposure. (b) PPA difference between groups based on pre- and post-exposure subtraction. Blue plots are for group A, orange for group B and green for group C. Data are in millimeters.

4.1.3. Stereoscopic Acuity. The Shapiro-Wilk test was performed to test the distribution of each group. No pre-exposure data follow a normal distribution. Data collected post-exposure for groups A and B follow a normal distribution, respectively p = 0.079 and p = 0.139, but not for groups C. In order to test the between groups' data, we subtracted the scores pre- and post-exposure, the results did not follow a normal distribution. The lower the TNO test score, the lower the stereoscopic acuity.

4.1.3.1.Difference pre-/post-exposure within each group. Stereoscopic acuity in group A fell only slightly by -8.03% on average between pre- and post-exposure as well as for group B by -4.50% but group C increased slightly by +1.6%, on average between pre- and post-exposure. The data are paired. A two-tailed Wilcoxon Sign Rank Test indicated for group A, that the median pre-exposure ranks, MD = -6.11, was statistically different from the median post-exposure ranks W = 4, **p = 0.024**. Group B (MD = 3.83, W = 3, p = 0.12) and C (MD = -1.86, W = 12, p = 0.792) tests Wilcoxon tests were not significant.

Stereoscopic acuity thus decreases between pre- and post-exposure for group A but no statistically significant result has been obtained for groups B and C.

4.1.3.2.Difference between groups. The data are independent. The Kruskal-Wallis Test indicated that there was no significant difference in stereoscopic acuity pre- and post-exposure between groups (H(2.79) = 2, p = 0.248), with a mean rank of 28.417 for group A, 32.818 for group B and 37.348 for group C.

As a conclusion, we did not observe statistically significant differences in stereoscopic acuity pre- and post-exposure in each group and between groups. Therefore, H3 cannot be supported via this optometric variable.

4.1.4. Visual acuity. The Shapiro-Wilk test was performed to test the distribution of each group pre- and post-exposure data as well as subtraction of the scores pre- and post-exposure for the between groups' comparison: none followed a normal distribution. The lower the test's score, the lower the visual acuity.

4.1.4.1.Difference pre-/post-exposure within each group. Visual acuity of group A fell by -10.83%, as well as group B by -8.96% and group C by -3.95%, on average between pre- and post-exposure (Fig. 5).

The data are paired. A Wilcoxon Sign Rank Test indicated for group A that the median pre-exposure ranks MD = 0.20, was statistically different from the median post-exposure ranks Z = -2.934, **p = 0.003**; for group B that the median pre-exposure ranks, MD = 0.40, was statistically different from the median post-exposure ranks Z = -2.223, **p = 0.026**; A two-tailed Wilcoxon Sign Rank Test indicated for group C, that the median pre-exposure ranks, MD = 0.31, was not statistically different from the median post-exposure ranks W = 3.5, p = 0.129. It can be seen that the visual acuity of group C fell only slightly by -4.89% on average between pre- and post-exposure.

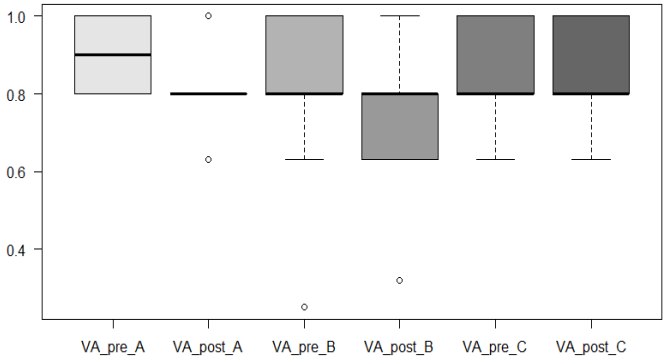

Figure 5: Visual Acuity pre- and post-exposure by group, lighter colors are for pre-exposure and darker colors are for post-exposure. Blue plots are for group A, orange for group B and green for group C. Data are in 10th corresponding to Scores at the E of Raskin Test

4.1.4.2.Difference between groups. The data are independent. The lower the score, the lower the visual acuity. The Kruskal-Wallis Test indicated that there was no significant difference in visual acuity pre- and post-exposure between groups (H(4.875 = 3, p = 0.181) with a mean rank of 25.917 for group A, 28,227 for group B and 34 for group C. Dunn's Post-Hoc tests without correction were conducted to obtain details of significance for each group compared. Only the difference between groups A and C is significant (MD = -0.55, SE = 0.25) Z = -1.860, **p = 0.031**. The difference between groups B and C (MD = -0.31, SE = 0.25) Z = -0.972, p = 0.165 and groups A and B (MD = -0.23, SE = 0.25) Z = -0.853, p = 0.197 are not significant.

As a conclusion, the median tendency indicates that condition A was the most tiring for visual acuity followed by condition B and finally C with significant differences for test groups A and B (pre- vs post-exposure) and between group A and group C. Therefore, H3 can be partially supported via this optometric variable.

4.2. Learning

4.2.1. Scores. The Shapiro-Wilk test was performed to test the distribution of each group. All G1 data follow a normal distribution (G1: A, p = 0.323; B, p = 0.083; C, p = 0.099) and only group A for G2 (p = 0.074). In order to test the between groups' data, we subtracted the scores G2 and G1, the results were tested with the Shapiro-Wilk test: all followed a normal distribution. The higher the in-game score, the higher the learning (memorization).

4.2.1.1.Difference pre-/post-exposure within each group. There is an improvement (+2.769%) between G1 and G2 for group A as well as for group B (+12.669%) and group C (+8.535%) although only group B showed a significant difference between in-game scores G1 and G2.

The data are paired. A single-samples t-test was conducted to compare subjects in game scores for the first game (G1) and the second game (G2). There was not a significant difference in the in-game scores for G1 (M = 37625, SD = 5852) and G2 (M = 38667, SD = 12021), t (23) = -0.494, p = 0.626 of the Group A. For group B, a Wilcoxon Sign Rank Test indicated that the median scores' G2 ranks, MD = 2285.71, was statistically different from the median G1 scores' ranks Z = -2.954, **p = 0.003**. For group C, a Wilcoxon Sign Rank Test indicated that the median scores' G2 ranks, MD = -4954.55, was not statistically different from the median G1 scores' ranks Z = -1.769, p = 0.079.

4.2.1.2.Difference between groups. The data are independent. A one-way between subjects ANOVA was conducted. An analysis of variance showed that the effect of the apparatus (PC or HMD) and displayed imaging (biocular or S3D) on in game score's evolution between G1 and G2 was not significant [F(2.66) = 1.181, p = 0.313].

The in-game scores G1 and G2 in the group B are statistically different. There is not a significant difference for the groups A and C. The null hypothesis can be rejected for the difference between G1 and G2 only for group B. The medians tendency indicates that condition B recorded better learning followed by condition C and finally condition A. Therefore, H1 can be partially supported via the score variable between the two games of group B. H4 can be supported because condition B shows statistically significant learning while condition C does not.

4.2.2. Time for answering. The Shapiro-Wilk test was performed to test the distribution of each group. All G1 and G2 data follow a normal distribution (G1: A, p = 0.265; B, p = 0.233; C, p = 0.637; G2: p = 0.289; B, p = 0.089; C, p = 0.053). In order to test the between groups' data, we subtracted the scores G2 and G1, the results were tested with the Shapiro-Wilk test: all followed a normal distribution.

4.2.2.1.Difference pre-/post-exposure within each group. The data are paired, the lower the score, the higher the learning. A single-samples t-test was conducted to compare subjects' time to answer for the first game (G1) and the second game (G2). There was a significant difference in subjects' time to answer for G1 (M = 12.72, SD = 3.24) and G2 (M = 8.44, SD = 2.19), t (23) = 7.35, **p = < 0.001** of the Group A. There was a significant difference in subjects' time for answering for G1 (M = 11.48, SD = 2.47) and G2 (M = 7.66, SD = 1.72), t (21) = 7.84, **p = < 0.001** of the Group B. There was a significant difference in subjects' time to answer for G1 (M = 7.99, SD = 1.29) and G2 (M = 5.23, SD = 1.35), t (22) = 12.72, **p = < 0.001** of the Group C. Subject's took less time to answer for the three conditions during the second game compared to the first one.

4.2.2.2.Difference between groups. The lower the score, the higher the learning. Group A took 4.281 minutes less to respond in G2 than in G1. Group B took 3.818 minutes less to respond in G2 than in G1. Group C took 2.763 minutes less to respond in G2 than in G1. The medians tendency indicates that participants in condition A recorded better learning followed by condition B and finally condition C (*Fig. 6*). A one-way between subjects ANOVA was conducted. An analysis of variance showed that the effect of the apparatus (PC or HMD) and displayed imaging (binocular or S3D) on times for answering difference between G1 and G2 was not significant [F(2, 66) = 2.91, p = 0.062]. Therefore, H1 can be partially supported via the evolution of response time variable between the first and second game. However, the difference in response time is not significantly different when conditions are tested one group with another. H4 can be supported because condition A and B show learning that tends to be higher than C although not statistically significant.

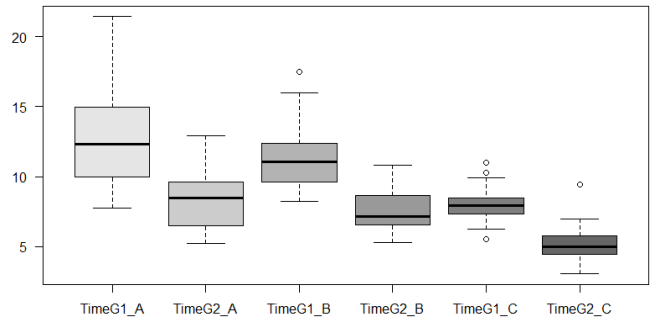

Figure 6: Time for answering G1 and G2 by group. Data are in minutes.

4.3. Quality of Experience

All questionnaires were completed after the two games (after the 30 minutes) and the post-exposure optometric measures.

4.3.1. Visual discomfort. All the 11 items from the questionnaire by Zeri and Livi were combined to test the visual comfort as a whole. The higher the score, the more subjects reported visual discomfort. A non-parametric Friedman test of differences among repeated measures was conducted and rendered a Chi-square value of 34.8 which was significant (**p < 0.001**). A Pairwise Durbin-Cornover post-hoc test was conducted and showed no significant difference in visual comfort for group A (M = 1.81, SD = 0.998) and group B (M = 1.76, SD = 0.933), p = 0.626; a significant difference in visual comfort for group A and group C (M = 1.38, SD = 0.688), **p = < 0.001**; a significant difference in visual comfort for group B and group C, **p = < 0.001**. Overall visual discomfort was reported as low by subjects, with medians ranging from 1 to 2 (Fig. 7). However, when groups are compared with one another, there is a significant difference in discomfort. The paired comparison of the groups shows that the difference is significant when control condition (C) is compared with test conditions (A and B). Therefore, the HMD groups reported greater visual discomfort. This is consistent with the results obtained with visual fatigue measurements and supports H2.

4.3.2. Presence. All 10 items from the Multimodal Presence Scale by Makransky *et al.* were combined to test the presence as a whole. The higher the score, the higher the level of presence reported. A non-parametric Friedman test of differences among repeated measures was conducted and rendered a Chi-square value of 3.80 which was

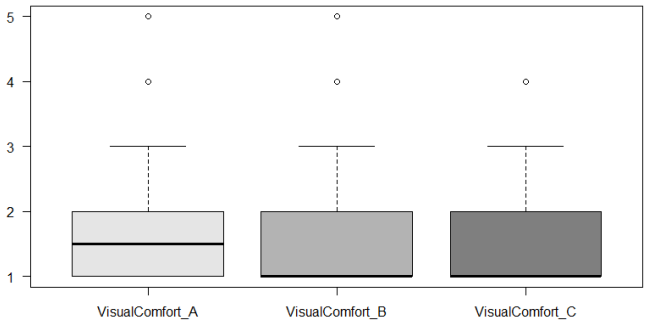

Figure 7: Visual discomfort questionnaire scores all items together (based on likert scales graduated 1-5)

not significant (p = 0.150). The medians for the three groups are identical (M = 3). The subjects therefore assessed their overall social and spatial presence similarly. Therefore, H2 can't be supported by this parameter.

4.3.3. Flow. All 9 items from the Flow Short Scale by Rheinberg, Vollmeyer and Engeser were combined to test the Flow as a whole. The higher the score, the higher the level of flow reported. A non-parametric Friedman test of differences among repeated measures was conducted and rendered a Chi-square value of 1.03, which was not significant (p = 0.598). The medians for the three groups are identical (M = 4). The subjects therefore assessed their overall flow similarly. Therefore, H2 can't be supported by this parameter.

5 DISCUSSION

5.1. Eyestrain

To summarize, we have evaluated eyestrain through several optometric measurements and showed that S3D drives to more eyestrain than biocular imaging (H3) based on punctum proximum of accommodation, stereoscopic acuity, ease of accommodation, but not based on visual acuity. This is consistent with previous studies [14, 48].

The Punctum Proximum of Accommodation is particularly impacted with our apparatus. The PPAs are significantly different pre- and post-exposure as they are also between conditions. The post-hoc comparison showed that the difference was significant between the HMDs conditions and the PC condition but not between S3D (group B) and the biocular imaging (group A). Yet, the tendencies of medians show that S3D drives to higher eyestrain based on PPA. The stereoscopic acuity gets impacted in a statistically significant way only for group A and with a two-tailed test. Therefore, we can't support the hypothesis based on that variable. The ease of accommodation can support the H3 hypothesis only based on medians' differences. The visual acuities pre- and post-exposure from the HMDs condition are significantly different but not when tested between conditions. Yet, the medians tendency shows that S3D drives to higher eyestrain based on visual acuity.

The difference between groups A and B not being statically significant can be explained by the fact that the in-depth objects of our scene were the text and boxes that the subjects had to select for answering. These boxes were not always on screen: the dialog scenes (when the friend and the recruiter are talking) did not show any objects in depth. Furthermore, the tasks in the serious game are not requiring moving the head too much and there are very few moving objects in the scene. Moreover, the vergence-accommodation conflict was less important which resulted in low visual stress. Furthermore, subjects adjusted HMDs and distance between screen and their eyes autonomously which means they might adjust it wrong for their eyes. Our results are consistent with previous work [29, 30, 36, 46] and show that HMDs have a more negative impact on the human visual system than computer screens when we compare the medians. Scenes and tasks asking for more movements and with more in-depth objects might lead to higher eyestrain in S3D conditions than binocular imaging conditions. Nevertheless, we can observe that eyestrain based on the PPA and visual acuity is higher with HMDs.

5.2. Learning efficiency

To summarize, we have evaluated learning curves through several in-game collected data and showed that learning expected answers in a job interview in SG 1) is not less efficient when the learners present eyestrain, which is not consistent with H1 and 2) is higher with VR-HMD than with a computer screen, which is consistent with H4.

That can be explained by the fact that interaction within the SG was almost similar among conditions. Concerning time for answering and eyestrain (H1), the three groups took statistically significant less time during their second game compared to their first game. But, the difference between groups is not statistically different. The medians tendency shows that the test groups (A and B) with HMDs had a better learning than the control group with PC. H1 can be partially supported. Learning expected answers in a job interview in SG is less efficient when the learners are presenting eyestrain based on time since group A performed better than group B, as shown in previous work. Learning expected answers in a job interview in a SG is higher with VR-HMD than with a computer screen.

Concerning comparison between VR-HMD and computer screen (H4), conditions A and B show better learning than C although not statistically significant. This is consistent with previous work [39, 49] about short-term retention. The results can be explained by the Human-Machine Interaction habits of subjects, which are more anchored with PCs than with HMDs. Therefore, subjects might be more efficient, from first game, with their interactions with the SG for the PC condition since they are experts with that apparatus, while the learning curve could be more important with HMD conditions. As for the learning efficiency linked to eyestrain, previous works showed that S3D leads to more eyestrain than *biocular* imaging, because of the vergence-accommodation conflict,

which can also drives to more cognitive fatigue [28] or be an indicator of it [16]. But S3D was not determinant in our SG to perform as shown by the results about the in-game scores since no learning was based on visuo-spatial detection or interaction unlike previous work [10, 23, 32].

5.3. Quality of experience

To summarize, we have evaluated the quality of experience through questionnaires and showed that eyestrain negatively impact quality of experience, supporting H2.

The subjects from the control group (C) reported statistically significant less discomfort than the two conditions with HMDs which is consistent with the tendencies observed with eyestrain and previous work [12] but unlike others [21]. There is no statistically significant difference between the HMD-S3D condition compared to the HMD-Biocular condition. Notwithstanding, H2 is supported since eyestrain deteriorates the quality of experience in a SG because subjects reported higher discomfort while eyestrain was also higher (with the A and B conditions). However, the Presence, unlike previous work [11, 41, 45], and Flow, in line with previous work [44], reported by subjects did not show any significant difference rather statistically or with medians. These results are not in line with previous work [15, 39]. This can be explained by the fact that our SG was the same on each device and did not have different instructions while previous work reported this variability of instructions as a possible explanation for better results with HMDs rather than with PCs [49]. Therefore, only visual discomfort supports our hypothesis about the negative impact of eyestrain over quality of experience.

5.4. Limitation of the study

The conditions tested in the present study have a high variability on oculo-motor demand since imaging type and devices displaying them are different. Therefore, the measured eyestrain to discuss its possible effects on learning is circumscribed to our apparatuses. Samsung Gear VR does not support the individualization of the inter-pupillary distance whereas this is possible in HTC Vive and Oculus Rift. Therefore, the tuning of Samsung Gear VR is less efficient and might influence oculomotor stress especially with stereoscopic content which we adapted based on general inter-pupillary distance. Our optometric results could vary depending on the Head-Mounted Display used.

6 CONCLUSION AND FUTURE WORK

When playing about 30 minutes of a SG to learn job interview, our study shows that subjects' learning curves is more efficient while using HMDs than a computer display. However, eyestrain tends to be higher while using HMDs. This eyestrain tends to have negative impacts on learning performances. The quality of experience is negatively impacted for visual comfort with HMDs compared to computer screen. Nevertheless, there are no significant difference between S3D and biocular imaging. Presence and Flow was similar among test and control conditions. Since our results show that eyestrain tends to be higher with S3D, not using S3D for SG with low visuo-spatial interactions and learning could be a way to avoid such limits and improve user experience quality. This solution would however not be satisfying with expert users or to learn certain tasks that are typically taking advantage of S3D such as surgical skills. Thus, finding a way to maintain S3D imaging by lowering its impact should also be investigated.

This study was conducted in an experimental environment as close to the real uses as possible. Lesser variability in conditions, more fundamental works, could help better understand links between eyestrain and learning. For example, by assessing visual fatigue and its impacts on learning in other experimental paradigms.

An analysis of learning retention scores after a few weeks may help to better understand the impact of HMDs, S3D and eyestrain on learning. Since S3D can be necessary depending on learning expectations, software solutions to reduce eyestrain should be explored.

ACKNOWLEDGMENTS

This work was supported by Manzalab Group (Paris, France), IDEFI-CréaTIC Programme: Initiatives of Excellence in Innovative Training and the French National Association of Research and Technology (ANRT) with the Grant CIFRE 2016-1571. The authors would like to thank Professor Ghislaine Azémard for her overall supervision of the study.

REFERENCES

[1] Alexander, A.L., Brunyé, T. and Sidman, J. 2005. From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *Aptima Inc. Woburn, MA*. (2005). DOI:https://doi.org/10.1037/e500852012-001.

[2] Banks, M.S., Allison, R.S. and Watt, S.J. 2012. Stereoscopy and the Human Visual System. *SMPTE Motion Imaging Journal*. 121, 4 (2012), 24–43. DOI:https://doi.org/10.5594/j18173.

[3] Bouvier, P., de Sorbier, F., Chaudeyrac, P. and Biri, V. 2008. Cross Benefits Between Virtual Reality And Games. In *Proceedings of the 1st. International Conference and Industry Symposium on Computer Games, Animation, Multimedia, IPTV, Edutainment and Security (CGAT'08)*. Singapore, 1-10. DOI:https://doi.org/10.5176/978-981-08-8227-3_cgat08-26

[4] Bowman, D.A., Sowndararajan, A., Ragan, E.D. and Kopper, R. 2009. Higher Levels of Immersion Improve Procedure Memorization Performance. (Lyon, France, 2009), 121–128.

[5] Broad, M.L. 1997. Overview of Transfer of Training: From Learning to Performance. *Performance Improvement Quarterly*. 10, 2 (1997), 7–21. DOI:https://doi.org/10.1111/j.1937-8327.1997.tb00046.x.

[6] Csikzentmihalyi, M. 1990. *Flow, the psychology of optimal experience*. Harper & Row, New York, NY.

[7] Cummings, J.J. and Bailenson, J.N. 2015. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*. 19, 2 (2015), 272–309. DOI:https://doi.org/10.1080/15213269.2015.1015740.

[8] Emoto, M., Niida, T. and Okano, F. 2005. Repeated Vergence Adaptation Causes the Decline of Visual Functions in Watching Stereoscopic Television. *Journal of Display Technology*. 1, 2 (2005), 328–340. DOI:https://doi.org/10.1109/jdt.2005.858938.

[9] Engeser, S. and Rheinberg, F. 2008. Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*. 32, 3 (2008), 158–172. DOI:https://doi.org/10.1007/s11031-008-9102-4.

[10] Fergo, C., Burcharth, J., Pommergaard, H.-C., Kildebro, N. and Rosenberg, J. 2017. Three-dimensional laparoscopy vs 2-dimensional laparoscopy with high-definition technology for abdominal surgery: a systematic review. *The American Journal of Surgery*. 213, 1 (2017), 159–170. DOI:https://doi.org/10.1016/j.amjsurg.2016.07.030.

[11] Freeman, J., Avons, S.E., Pearson, D.E. and IJsselsteijn, W.A. 1999. Effects of Sensory Information and Prior Experience on Direct Subjective Ratings of Presence. *Presence: Teleoperators and Virtual Environments*. 8, 1 (1999), 1–13. DOI:https://doi.org/10.1162/105474699566017.

[12] Guo, J., Weng, D., Duh, H.B.-L., Liu, Y. and Wang, Y. 2017. Effects of using HMDs on visual fatigue in virtual environments. In *Proceedings of the 2017 IEEE Virtual Reality (VR)*. Los Angeles, CA, USA, 249-250. DOI: https://doi.org/10.1109/VR.2017.7892270.

[13] Hamari, J., Shernoff, D.J., Rowe, E., Coller, B., Asbell-Clarke, J. and Edwards, T. 2016. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*. 54, (2016), 170–179. DOI:https://doi.org/10.1016/j.chb.2015.07.045.

[14] Hoffman, D.M., Girshick, A.R., Akeley, K. and Banks, M.S. 2008. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*. 8, 3:33 (2008), 1–30. DOI:https://doi.org/10.1167/8.3.33.

[15] Hupont, I., Gracia, J., Sanagustin, L. and Gracia, M.A. 2015. How do new visual immersive systems influence gaming QoE? A use case of serious gaming with Oculus Rift. In *Proceedings of the 7th. International Workshop on Quality of Multimedia Experience (QoMEX)*. Costa Navarino, Messinia, Greece, 1–6. DOI:https://doi.org/10.1109/QoMEX.2015.7148110.

[16]Iwasaki, T. 1993. Effects of a visual task with cognitive demand on dynamic and steady-state accommodation. *Ophthalmic and Physiological Optics*. 13, 3 (1993), 285–290. DOI:https://doi.org/10.1111/j.1475-1313.1993.tb00470.x.

[17]Karajeh, H., Maqableh, M. and Masa'deh, R. 2014. A Review on Stereoscopic 3D: Home Entertainment for the Twenty First Century. *3D Research*. 5, 4 (2014). DOI:https://doi.org/10.1007/s13319-014-0026-3.

[18] Joseph R. Keebler. 2011. *Effects of 3D stereoscopy, visuo-spatial working memory, and perceptions of simulation experience on the memorization of confusable objetcs*. Ph.D. Dissertation. University of Central Florida, Orlando, FL. UMI Order Number: CFE0003939.

[19] Keller, K. and Colucci, D. 1998. Perception in HMDs: what is it in head-mounted displays (HMDs) that really make them all so terrible? In *Proceedings of the SPIE 3362, Helmet- and Head-Mounted Displays III*. Orlando, FL, 46–53. DOI: https://doi.org/10.1117/12.317454.

[20] Kozulin, P., Ames, S.L. and McBrien, N.A. 2009. Effects of a Head-Mounted Display on the Oculomotor System of Children. *Optometry and Vision Science*. 86, 7 (2009), 845–856. DOI:https://doi.org/10.1097/opx.0b013e3181adff42.

[21] Kulshreshth, A., Schild, J. and LaViola, J.J. 2012. Evaluating user performance in 3D stereo and motion enabled video games. In the *Proceedings of the International Conference on the Foundations of Digital Games (FDG'12)*. Raleigh, NC, 33–40. DOI:https://doi.org/10.1145/2282338.2282350.

[22] Limperos, A., Waddell, T.F., Ivory, A.H. and Ivory, J.D. 2014. Psychological and physiological responses to stereoscopic 3d presentation in handheld digital gaming: Comparing the experiences of frequent and infrequent game. *Presence: Teleoperators and Virtual Environments*. 23, 4 (2014), 341–353. DOI:https://doi.org/10.1162/pres_a_00204.

[23] Loup-Escande, E., Jamet, E., Ragot, M., Erhel, S. and Michinov, N. 2017. Effects of Stereoscopic Display on Learning and User Experience in an Educational Virtual Environment. *International Journal of Human-Computer Interaction*. 33, 2 (2017), 115–122. DOI:https://doi.org/10.1080/10447318.2016.1220105.

[24] Morse, S.E. and Jiang, B.-C. 1999. Oculomotor Function After Virtual Reality Use Differentiates Symptomatic from Asymptomatic Individuals. *Optometry and Vision Science*. 76, 9 (1999), 637–642. DOI:https://doi.org/10.1097/00006324-199909000-00021.

[25] Makransky, G., Lilleholt, L. and Aaby, A. 2017. Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*. 72, (2017), 276–285. DOI:https://doi.org/10.1016/j.chb.2017.02.066.

[26] McIntire, J.P., Havig, P.R. and Geiselman, E.E. 2012. What is 3D good for? A review of human performance on stereoscopic 3D displays. In *Proceedings of SPIE 8383, Head- and Helmet-Mounted Displays XVII; and Display Technologies and Applications for Defense, Security, and Avionics VI, 83830X*. Baltimore, MD, 1-13. DOI:https://doi.org/10.1117/12.920017.

[27] McMahan, A. 2003. Immersion, engagement and presence. *The Oxford Handbook of Interactive Audio*. K. Collins, B. Kapralos, and H. Tessler, eds. Oxford University Press. 67–86.

[28] Mistry, M., Roach, V.A. and Wilson, T.D. 2013. Application of Stereoscopic Visualization on Surgical Skill Acquisition in Novices. *Journal of Surgical Education*. 70, 5 (2013), 563–570. DOI:https://doi.org/10.1016/j.jsurg.2013.04.006.

[29]Mon-Williams, M. and Wann, J.P. 1998. Binocular Virtual Reality Displays: When Problems Do and Don't Occur. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 40, 1 (1998), 42–49. DOI:https://doi.org/10.1518/001872098779480622.

[30] Mon-Williams, M., Warm, J.P. and Rushton, S. 1993. Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display. *Ophthalmic and Physiological Optics*. 13, 4 (1993), 387–391. DOI:https://doi.org/10.1111/j.1475-1313.1993.tb00496.x.

[31] Mukai, A., Yamagishi, Y. and Hirayama, M.J. 2011. Effects of stereoscopic 3D contents on the process of learning to build a handmade PC. *Knowledge Management and E-Learning*. 3, 2 (2011), 491–506. DOI:https://doi.org/10.5392/jkca.2010.10.2.138.

[32] Price, C.A., Lee, H.-S., Subbarao, M., Kasal, E. and Aguilera, J. 2015. Comparing Short- and Long-Term Learning Effects Between Stereoscopic and Two-Dimensional Film at a Planetarium. *Science Education*. 99, 6 (2015), 1118–1142. DOI:https://doi.org/10.1002/sce.21185.

[33] Patterson, R., Winterbottom, M.D. and Pierce, B.J. 2006. Perceptual issues in the use of head-mounted visual displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 48, 3 (2006), 555–573. DOI:https://doi.org/10.1518/001872006778606877.

[34] Pavlas, D., Heyne, K., Bedwell, W., Lazzara, E. and Salas, E. 2010. Game-based learning: The impact of flow state and videogame self-efficacy. In *Proceedings of the 54th. Human Factors and Ergonomics Society Annual Meeting*. San Francisco, CA, 2398–2402. DOI:https://doi.org/10.1177/154193121005402808.

[35] Price, A. and Lee, H.-S. 2010. The Effect of Two-dimensional and Stereoscopic Presentation on Middle School Students' Performance of Spatial Cognition Tasks. *Journal of Science Education and Technology*. 19, 1 (2010), 90–103. DOI:https://doi.org/10.1007/s10956-009-9182-2.

[36] Rushton, S.K. and Riddell, P.M. 1999. Developing visual systems and exposure to virtual reality and stereo displays: some concerns and speculations about the demands on accommodation and vergence. *Applied Ergonomics*. 30, 1 (1999), 69–78. DOI:https://doi.org/10.1016/s0003-6870(98)00044-1.

[37] Sanchez-Vives, M.V. and Slater, M. 2005. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*. 6, 4 (2005), 332–339. DOI:https://doi.org/10.1038/nrn1651.

[38] Santos, B.S., Dias, P., Pimentel, A., Baggerman, J.-W., Ferreira, C., Silva, S. and Madeira, J. 2008. Head-mounted display versus desktop for 3D navigation in virtual reality: a user study. *Multimedia Tools and Applications*. 41, 1 (2008), 161–181. DOI:https://doi.org/10.1007/s11042-008-0223-2.

[39] Schmoll, L., Veit, M. and Roy, M. 2013. Serious game et apprentissage en réalité virtuelle: résultats d'une étude préliminaire sur la mémorisation en langue étrangère. *Synergies de pays germanophones 7*. (2013). DOI:https://doi.org/10.3726/978-3-0351-0935-1/6.

[40] Schor, C.M. 1992. A Dynamic Model of Cross-Coupling Between Accommodation and Convergence: Simulations of Step and Frequency Responses. *Optometry and Vision Science*. 69, 4 (1992), 258–269. DOI:https://doi.org/10.1097/00006324-199204000-00002.

[41] Schrader, C. and Bastiaens, T.J. 2012. The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior*. 28, 2 (2012), 648–658. DOI:https://doi.org/10.1016/j.chb.2011.11.011.

[42] Slater, M. and Wilbur, S. 1997. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*. 6, 6 (1997), 603–616. DOI:https://doi.org/10.1162/pres.1997.6.6.603.

[43] Stone, R.J., Panfilov, P.B. and Shukshunov, V.E. 2011. Evolution of aerospace simulation: From immersive Virtual Reality to serious games. In P*roceedings of 5th. International Conference on Recent Advances in Space Technologies - RAST2011*. Istanbul, Turkey, 655–662. DOI:https://doi.org/10.1109/RAST.2011.5966921.

[44] Takatalo, J., Kawai, T., Kaistinen, J., Nyman, G. and Häkkinen, J. 2011. User Experience in 3D Stereoscopic Games. *Media Psychology*. 14, 4 (2011), 387–414. DOI:https://doi.org/10.1080/15213269.2011.620538.

[45] Thalmann, D., Lee, J. and Thalmann, N.M. 2016. An evaluation of spatial presence, social presence, and interactions with various 3D displays. In *Proceedings of the 29th. International Conference on Computer Animation and Social Agents*. Geneva, Switzerland, 197–204. DOI:https://doi.org/10.1145/2915926.2915954.

[46] Ujike, H. and Watanabe, H. 2015. 3D visual fatigue: Interaction between ranges of binocular disparity and interocular crosstalk. In *Proceedings of the IEEE 4th. Global Conference on Consumer Electronics (GCCE)*. Osaka, Japan, 668-669. DOI:https://doi.org/10.1109/GCCE.2015.7398697.

[47] Ukai, K. and Howarth, P.A. 2008. Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays*. 29, 2 (2008), 106–116. DOI:https://doi.org/10.1016/j.displa.2007.09.004.

[48] Wann, J.P., Rushton, S. and Mon-Williams, M. 1995. Natural problems for stereoscopic depth perception in virtual environments. *Vision Research*. 35, 19 (1995), 2731–2736. DOI:https://doi.org/10.1016/0042-6989(95)00018-u.

[49] Webster, R. 2015. Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments*. 24, 6 (2015), 1319–1333. DOI:https://doi.org/10.1080/10494820.2014.994533.

[50] Zeri, F. and Livi, S. 2015. Visual discomfort while watching stereoscopic three-dimensional movies at the cinema. *Ophthalmic and Physiological Optics*. 35, 3 (2015), 271–282. DOI:https://doi.org/10.1111/opo.12194.