Analysing the impact of interaction, embodied interaction and head-mounted immersion on memorisation: rationale and experiment design

Jack Ratcliffe
Computer Science
Queen Mary University of London
London, UK
j.ratcliffe@qmul.ac.uk

Dr. Laurissa Tokarchuk
Computer Science
Queen Mary University of London
London, UK
l.tokarchuk@qmul.ac.uk

***Abstract*—Immersive technologies have proven effective for learning, particularly in simulated environments. Part of the success of simulation learning has been attributed to learners' ability to interact - in an embodied fashion - with objects in an immersive virtual environment. However, there are limited explorations that directly compare the composite features that have been credited for a simulation's learning successes: interaction vs. non-interaction; embodied controllers vs. abstracted control; and head-mounted displays vs. monitor-representation. This paper outlines a design that could be used to determine the extent to which each of these immersive features, and combinations of these, contribute to learning inside a single environment, allowing for more effective comparisons.**

Keywords— Immersion, Learning, Embodied Interaction, Head-Mounted Display, Language Learning

It is widely acknowledged that immersive computing environments, such as simulations, can successfully enable training and learning [1,2], and that this can transfer from the screen into real-world situations [3,4]. These systems typically leverage multiple immersive technologies, including interactive environments, immersive environments, and embodied interaction, in the belief that greater immersion leads to greater learning outcomes [2]. However, recent meta-analysis suggests that this view is potentially too simplistic. When the impact of individual immersive technologies on learning has been studied, some devices, such as interaction and head-mounted displays, appear to have a much greater or more consistent impact on learning than others, such as embodied control [5].

The different subjects and design goal discrepancies in existing studies make empirical meta-analysis difficult, however. The breadth of explored subjects, each prioritising different aspects suited to their content and previous pedagogical approaches, introduce many unique variables. There are important differences between learning environments designed to for cognition alongside accurate motor-skill control for gesture intensive operations (such as surgery) or reading emotional cues (such as nursing). While cognitive learning could happen in either of these examples, learners in the former would likely benefit from accurate embodied controls due to the nature of the material, while those in latter would likely benefit more from an enhanced sense of presence.

Acknowledging the role that subject-specificity plays in virtual reality learning interventions, and thus how each topic may have its immersive benefits and weaknesses, we, like Howard [5], still believe it is important to investigate the relationship between hardware and software and general VR intervention success. The central question - what is the effect of commonly applied hardware and software on VR intervention outcomes? - requires contrastive experimental design.

This paper suggests a design to clarify our understanding of the effects of three prominent immersive techniques. It outlines a study design that can explore whether interacting with virtual objects creates a powerful learning effect, whether that effect is more pronounced when the interaction is embodied, and the extent to which using an immersive head-mounted display enhances those effects over a traditional monitor display.

I. BACKGROUND

Perhaps the most-explored, highly immersive learning comes from the simulation space, which has decades of academic pedigree and real-world industry applications. Flight simulators and surgical simulations have repeatedly proved that immersive education can successfully inform learners at a rate comparable to that of real-world experience, and that this learning is transferable to real-world applications (see surgical meta-analysis by Alaker [6] or Haque [7]).

While many older simulations used augmented physical tools as part of the simulation (e.g. an enclosed cab for pilots or a laparoscopy for surgeons), more recent explorations in flight and surgery simulation which leverage generic immersive technologies (e.g. the Oculus Rift head-mounted display and Leap Motion hand-detection controller) show that effective learning outcomes still occur with virtual tool representation [5, 8, 9].

If the success of immersive learning is not dependent on having access to the exact tools (evidenced by countless studies that do not use exact tools, or are entirely cognitive), then it is reasonable to extend our exploration into other areas: the ability to interact with a scenario; the embodiment of that interaction; or the immersive environment in which the scenario is presented.

A. Immersive Environment

It is widely accepted that being highly engaged with a computer learning system provides beneficial learning outcomes. It is also well-established that the use of immersive technologies that prioritise user tracking level, stereoscopic vision, and field of view (such as head-mounted displays), allow users to feel so engaged in an environment that they feel physically present in the virtual environment [10]. This presence is then theorised to increase the effectiveness of the mediated environment [11], as learners may be more mentally engaged with the presented information [12, 13]. Howard's meta-analysis found that HMDs and other types of immersive output had a significant effect for learning outcomes. However, there is also strong evidence countering this conclusion. For flight simulators, various studies [14, 15] demonstrated home-PC software experienced on a monitor had equivalent benefits to the much more immersive enclosed cab simulator environment. Studies comparing learning inside HMDs with monitor often report enhanced presence and engagement, but little difference in the actual learning outcomes [16]. Some research even evidences an inverse link between being immersed in an environment and learning due to increased cognitive load demands [17]. It is possible that general learning benefits from HMDs come from their ability to reduce outside distractions, as the user's entire field of vision becomes the display [18].

Immersive environment learning success has more often been linked with the affective factors surrounding learning, such as emotional preparation [19]; motivation and enjoyment [17]; or anxiety reduction [20]. These aspects are not to be dismissed, as motivation and enjoyment are key factors in long-term educational engagement and positive learning outcomes [21], but they are not evidence that immersive environments provide better immediate knowledge-transfer. This is an important distinction if we are designing for one-off training rather than long-term curricular engagement.

There is an argument to suggest that immersive environments could play an enhanced role in language education compared with other subject areas, as being immersed in a relevant learning context is considered important in second language acquisition. Language teaching environments that are closer to the original socio-culture of the target language, or can replicate aspects of it, are believed to have better learning outcomes [22]. Evidence for this exists in positive analysis of non-virtual immersion learning through study-abroad programmes [23]. Therefore, by using language memorisation as a subject for exploring the efficacy of immersive environment learning, we can examine whether the immersion has an effect in a theoretically-optimal setting.

B. *Embodiment & Interaction*

It is widely understood that rich interaction with objects in an environment, whether embodied or not, has a significant impact on learning compared with not being able to interact [24, 25] and that interaction (amongst other game-elements) can improve cognition in VR [26]. However, it is currently unclear whether embodied interaction - enabled by inputs such as gestural controllers or hand and body tracking - is significantly most conducive to learning than non-immersive interaction. There is evidence for its added efficacy [27, 28] but also against any benefits compared with normal interaction ("SpatialEase" [29]). Similarly, desktop-based flight simulators, interacted with mouse and keyboard or joystick, have proven the match for full cockpit replicas [14, 15]. Howard's meta-analysis found that embodied interaction has little impact on learner outcomes [5].

Many other investigations in this area are of limited value for this discussion, as interactional richness is often added to their embodied controls but restricted from their nonembodied interfaces, such as comparing the dynamic creation of 3D models with relatively dry text entry [30].

Similar to immersive environment studies, embodied learning studies record many positive affective factors, mostly relating to student motivation [31, 32, 33]. However, as mentioned previously, while motivation strongly affects learning, it does not mean that embodied learning is itself a more effective learning process, especially for short-term engagements.

C. *Multimodal relationships*

Generally, immersive input and output hardware have been studied in isolation [5]. Investigations into the multimodal relationships between interaction, embodied interaction and HMD-aided immersion are limited. Vázquez's [34] study of verb acquisition demonstrated that embodied interaction inside an HMD provides better learning and retention than learning without interaction. Unfortunately, it does not explore whether the embodiment is a key factor in the learning process, or the interaction.

Howard recommends for more research to study the effects of immersive types together, as they are often "developed from a common theoretical perspective" and that it is possible that "interaction effects may be present between many specialized input hardware and advanced output hardware, and these interactions are undiscovered in current research".

From the literature, it is clear that we need to develop a greater understanding of the impact of different immersive technologies on learning, their strengths and drawbacks, and whether they work in harmony or conflict in an education setting. It is not enough to, as many do, simply assume that greater immersion leads to greater learning outcomes, even in fields such as language acquisition, where immersion is considered a strong contributing part of the learning experience in the real world.

II. STUDY DESIGN

We propose a study designed to examine some of the questions that emerged from the above literature:

- Does interaction lead to better memorisation than non-interaction?
- Does embodied interaction lead to better memorisation than non-embodied interaction?
- Does an HMD with embodied interaction lead to better memorisation than a monitor with embodied interaction?
- Is there any difference between memorisation of the action, and the object of the action?

A. *Procedure*

Participants will be exposed to the learning environment via a monitor, followed by the same environment with a head-mounted display (or vice versa). Inside the environment are three interaction types: objects that the participant can interact with using embodied controls, objects they can interact with abstractedly ("push X button to do Y") and objects they cannot interact with, which depict an action (see Table 1). Each action and object are a noun/verb pair; the noun corresponding to the object, and the verb corresponding to the participant's action, the participant's avatar's abstracted action, or a depiction of an action.

Participants will be presented with each object/action pair and their corresponding noun/verb pairing once, and then will be allowed ten minutes of free practice in the environment. We believe this presents a more natural self-learning experience than restricting their exposure to each object/pair a certain number of times or for a limited duration. The number of engagements with each object/action will be recorded.

Participants will be tested immediately prior to the experiment to gauge their base understanding, after each training session, and exactly one week after the experiment.

TABLE I. IMMERSION VARIABLES EXPERIENCED BY PARTICIPANTS

Interaction	Input	Output
No	Analogue stick	Desktop monitor
Yes	Analogue stick	Desktop monitor
Yes	Embodied control	Desktop monitor
No	Analogue stick	Head-mounted display
Yes	Analogue stick	Head-mounted display
Yes	Embodied control	Head-mounted display

B. *Corpus*

Participants will be tested on their knowledge of 30 noun/verb pairs, although only 24 noun/verb contextually-relevant pairs will be included in the learning environment. Of these, 12 pairs will be assigned to each monitor and HMD environment for each participant. These 12 pairs will then be distributed into three groups of five, one for each type of object interaction.

Word pairs will be distributed so that they appear an equal number of times for each interaction type, for each display type, and for the session in which they are presented (first or second).

The words are in Japanese and 外来語 (gairaigo; import words) have been removed to reduce the chance of subjects' inferring a meaning. The linguistic distance from European languages should ensure that prior or inferred knowledge is reduced for most UK-recruited participants. The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

C. *Inclusion and Exclusion Criteria*

Participants will be pre-tested to determine their existing proficiency with Japanese. The participant pool will consist of participants who self-report as having no prior knowledge or only "extremely basic" Japanese knowledge. Of there, a further pre-test evaluation will determine any familiarity with words used in the study. If a participant has an awareness of the meaning of any of the words they will be excluded from the study.

D. *Environment*

The setting is a first-person, Unity-based 3D environment with a realistic depiction of a coffee shop, with objects and actions that are contextually appropriate. We believe the use a of high context environment is more likely to determine any differences between monitor vs. HMD learning than a context-free experience, and it is unrealistic to imagine that any real-world learning application would exist in an entirely abstracted setting.

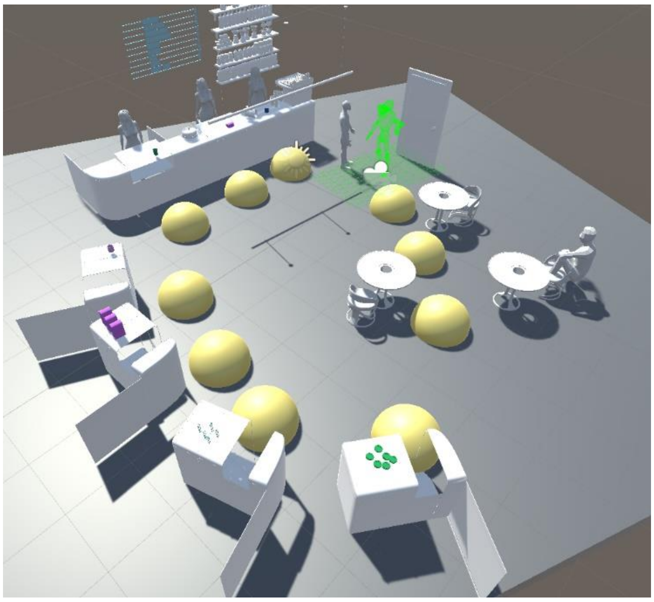

Fig. 1. Enviroment with textures removed, showing interaction zone markers (which are hidden from users). User avatar highlighted in green.

The environment includes a non-player animated pedagogical agent that guides participants through the experience. Non-player pedagogical agents have proven more conducive to learning than text-based explanations [35] and are a good-fit for demonstrating novel gestures. The agent will explain the noun to be learned, indicate the game object that relates to the noun (a realistic model of the real-world object), explain the verb to be learned, and enact the gesture corresponding to the verb. When a participant successfully performs the correct gesture with the noun, the interaction with that object is ended and the participant is asked to move to the next highlighted learning area. After all areas have been experienced, the participant can freely explore and repeat the learning exercises.

Participants will use either an Oculus Rift (HMD) with Touch controllers or a desktop monitor with Touch controllers. For both display types participants will use the Touch controllers' analogue sticks to move their avatar from a first-person perspective, using a combination of embodied actions or the trigger button to control the interaction types.

E. *Evaluation*

In pre- and post-exposure evaluations, participants will be asked to provide the English meaning of a Japanese noun or verb. Nouns and verbs will be tested separately to reduce the possibility for subjects to infer the meaning of one or the other. The evaluatory tests presented post-exposure to an environment consist of 15 nouns and 15 verbs, including three noun/verb pairs that the subject has not previously heard. The test presented one-week post-exposure includes 30 nouns and verbs, including six noun/verb pairs the subject has not previously heard. Participants will be instructed to leave an answer blank if they did not know it

Participants will be asked to complete the IPQ questionnaire (36) to provide insight on their feelings of spatial presence, involvement and experienced realism after each HMD and monitor engagement to enable comparative analysis between presence, output technology and learning outcomes. Participants will also be asked to take a VARK survey (37) to determine their learning preference to determine if a predominantly kinesthetic learning preference influences the outcome of interaction type.

F. Analysis

We will compare vocabulary recognition success between the different interaction types (interaction vs embodied interaction vs non-interaction), the different display systems (HMD vs monitor) and between both. We believe this will allow us to answer whether there is any difference in memorisation from the interaction types, between the display types, and whether multimodal immersion has an impact, and to what extent.

III. CONCLUSION

We outlined three immersive paradigms that are considered to have an impact on learning in immersive systems, despite mixed reports of their varying efficacies. Due to the lack of clarity regarding how these immersive elements affect learning, we presented a study designed to allow for direct comparisons between these elements and their effect on memorisation of language as a subset of general learning. We also noted limitations of this study, including the varying immersive requirements of different, subject-specific learning, and noted how some of the unique factors of language learning might reinforce the impact of presence and HMDs on memorisation in this subject.

REFERENCES

[1] G Burdea, Grigore C., and Philippe Coiffet. Virtual reality technology. John Wiley & Sons, (2003)

[2] Dede, Chris. "Immersive interfaces for engagement and learning." science 323.5910 (2009): 66-69

[3] Seymour, Neal E., et al. "Virtual reality training improves operating room performance: results of a randomized, double-blinded study." Annals of surgery 236.4 (2002): 458.

[4] Shin, Sujin, Jin-Hwa Park, and Jung-Hee Kim. "Effectiveness of patient simulation in nursing education: meta-analysis." Nurse education today 35.1 (2015): 176-182.

[5] Howard, Matt C. "Virtual reality interventions for personal development: A meta-analysis of hardware and software." Human–Computer Interaction 34.3 (2019): 205-239.

[6] Alaker, Medhat, Greg R. Wynn, and Tan Arulampalam. "Virtual reality training in laparoscopic surgery: a systematic review & meta-analysis." International Journal of Surgery 29 (2016): 85-94.

[7] Haque, Syed, and Shankar Srinivasan. "A meta-analysis of the training effectiveness of virtual reality surgical simulators." IEEE Transactions on Information Technology in Biomedicine 10.1 (2006): 51-58.

[8] Pulijala, Y., et al. "An innovative virtual reality training tool for orthognathic surgery." International journal of oral and maxillofacial surgery 47.9 (2018): 1199-1205.

[9] Barré, Jessy, et al. "Virtual Reality Single-Port Sleeve Gastrectomy Training Decreases Physical and Mental Workload in Novice Surgeons: An Exploratory Study." Obesity surgery (2019): 1-8.

[10] Cummings, James J., and Jeremy N. Bailenson. "How immersive is enough? A meta-analysis of the effect of immersive technology on user presence." Media Psychology 19.2 (2016): 272-309.

[11] Monahan, Teresa, Gavin McArdle, and Michela Bertolotto. "Virtual reality for collaborative e-learning." Computers & Education 50.4 (2008): 1339-1353.

[12] Bulu, Saniye Tugba. "Place presence, social presence, co-presence, and satisfaction in virtual worlds." Computers & Education 58.1 (2012): 154-161.

[13] Ling, Yun, et al. "The relationship between individual characteristics and experienced presence." Computers in Human Behavior 29.4 (2013): 1519-1530.

[14] Phillips, Sybil I., Charles L. Hulin, and Paul J. Lamermayer. "Uses of part-task trainers in instrument flight training." International Symposium on Aviation Psychology, 7 th, Columbus, OH. 1993.

[15] Ortiz, Gustavo A. "Effectiveness of PC-based flight simulation." The International Journal of Aviation Psychology 4.3 (1994): 285-291.

[16] Buttussi, Fabio, and Luca Chittaro. "Effects of different types of virtual reality display on presence and learning in a safety training scenario." IEEE transactions on visualization and computer graphics 24.2 (2018): 1063-1076.

[17] Makransky, Guido, Thomas S. Terkildsen, and Richard E. Mayer. "Adding immersive virtual reality to a science lab simulation causes more presence but less learning." Learning and Instruction (2017).

[18] Robinett, Warren, and Jannick P. Rolland. "A computational model for the stereoscopic optics of a head-mounted display." Presence: Teleoperators & Virtual Environments 1.1 (1992): 45-62.

[19] Ochs, Magalie, et al. "Training doctors' social skills to break bad news: evaluation of the impact of virtual environment displays on the sense of presence." Journal on Multimodal User Interfaces 13.1 (2019): 41-51.

[20] Powers, Mark B., and Paul MG Emmelkamp. "Virtual reality exposure therapy for anxiety disorders: A meta-analysis." Journal of anxiety disorders 22.3 (2008): 561-569.

[21] Kapp, Karl M. The gamification of learning and instruction. San Francisco: Wiley, 2012.

[22] Celce-Murcia, Marianne. "Rethinking the role of communicative competence in language teaching." Intercultural language use and language learning. Springer, Dordrecht, 2008. 4157.

[23] Kinginger, Celeste. "Language learning in study abroad: Case studies of Americans in France." The Modern Language Journal 92 (2008): i-131.

[24] Bedwell, Wendy L., et al. "Toward a taxonomy linking game attributes to learning: An empirical study." Simulation & Gaming 43.6 (2012): 729-760.

[25] Li, Ming-Chaun, and Chin-Chung Tsai. "Game-based learning in science education: A review of relevant research." Journal of Science Education and Technology 22.6 (2013): 877898.

[26] Hou, Jinghui, et al. "Effects of screen size, viewing angle, and players' immersion tendencies on game experience." Computers in Human Behavior 28.2 (2012): 617-623.

[27] Lindgren, Robb, et al. "Enhancing learning and engagement through embodied interaction within a mixed reality simulation." Computers & Education 95 (2016): 174-187.

[28] Chang, Yao-Jen, Shu-Fang Chen, and Jun-Da Huang. "A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities." Research in developmental disabilities 32.6 (2011): 2566-2570.

[29] Edge, Darren, Kai-Yin Cheng, and Michael Whitney. "SpatialEase: learning language through body motion." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2013.

[30] Johnson-Glenberg, Mina C., and Colleen Megowan-Romanowicz. "Embodied science and mixed reality: How gesture and motion capture affect physics education." Cognitive Research: Principles and Implications 2.1 (2017): 24.

[31] Anastopoulou, Stamatina, Mike Sharples, and Chris Baber. "An evaluation of multimodal interactions with technology while learning science concepts." British Journal of Educational Technology 42.2 (2011): 266-290.

[32] Dunleavy, Matt, Chris Dede, and Rebecca Mitchell. "Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning." Journal of science Education and Technology 18.1 (2009): 7-22.

[33] Jones, M. Gail, et al. "Haptic augmentation of science instruction: Does touch matter?." Science Education 90.1 (2006): 111-123.

[34] Christian Vázquez, Lei Xia, Takako Aikawa, Pattie Maes "Words in Motion: Kinesthetic Language Learning in Virtual Reality

[35] Atkinson, Robert K. "Optimizing learning from examples using animated pedagogical agents." Journal of Educational Psychology 94.2 (2002): 416.

[36] Schubert, T. "The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness." Zeitschrift fr Medienpsychologie, 15 (2003), 69-71.

[37] Fleming, N.D. & Mills, C. "Not Another Inventory, Rather a Catalyst for Reflection." To Improve the Academy, 11 (1992) 137-155