Emerging Challenges of Evaluating Immersive Virtual Reality Safety Training: An Early Design Science Perspective

Amir Haj-Bolouri[1][0000-0002-1202-9797] and Matti Rossi[2][0000-0001-8913-3747]

[1]School of Economics, Business, and IT, University West, Sweden
amir.haj-bolouri@hv.se

[2]Information Systems, Aalto University, Finland
matti.rossi@aalto.fi

Abstract:

In this research in progress paper, we outline the emerging challenges of evaluating Immersive Virtual Reality (IVR) training environments. The challenges were identified through an ongoing Action Design Research project where safety training is organized and conducted in IVR. The purpose of this on-going research was to study the evaluation forms of training in IVR environments. Thus, the question "What challenges for evaluation of design and learning outcomes do emerge in IVR-environments?" was asked for guidance and research. Group interviews and observations were conducted together with stakeholders and users of IVR-training. As a result, this research in progress paper reports early findings in forms of emerging challenges of evaluating IVR training and outlines steps for future research.

Keywords:
Immersive Virtual Reality, Training, Design Science, Evaluation

1 Introduction

The re-emergence of Immersive Virtual Reality (IVR) technologies in industrial sectors (e.g., mining, manufacturing, construction, transportation), especially facilitating education and training at work [1-3] has spurred interest into effective design of IVR environments. Companies employ IVR technologies such as Oculus Rift and HTC Vive [4] to mediate training scenarios that enhance employees' conceptual and representational knowledge [5], whereas other initiatives emphasize meaningful simulations that have a direct impact on employees' competencies at work (e.g., skills) [6]. An important component of VR training is the evaluation of employees' experiences and performance in relation to the training objectives [7].

Studies within Educational Science [8-9] and Safety Science [10-11] stress the importance of employing evaluation strategies that allow the researchers to systematically evaluate the learning outcomes (e.g., increased skills, competencies, factual knowledge, procedural knowledge) of IVR-training, rather than solely the usability of VR apps. At the same time, the design goals of IVR training environments are important to evaluate because the user interacts and learns from these environments [12]. Future research agendas for VR and training have explicitly been formulated with a specific attention to the evaluation part: *"Evaluations of educational VR applications need to be conducted both in terms of feasibility (i.e., from a software engineering standpoint) and of the learning outcomes (i.e., from a pedagogical standpoint). […] Thus, future research needs to include workshops, surveys, and focus group discussions in order to extract the necessary learning content and the expected learning outcomes as well as the usability requirements for VR applications from teachers and students."* [13, p. 22]

To find solutions to the abovementioned issues this research in progress asks the following research question: *What challenges for evaluation of design and learning outcomes do emerge in IVR-environments?* We are fully aware that we cannot answer the research question in its entirety through this research in progress, but rather we outline how this can be studied through early findings of an Action Design Research (ADR) [14] case of IVR safety training. We propose tentative outcomes from the case in the form of emerging challenges of evaluating IVR training.

The rest of this paper is structured as follows: first, we discuss the status on VR research and evaluation of training. After that, we present the ADR case. Then, we propose the identified challenges of evaluating IVR training. Finally, we end this paper through concluding remarks.

2 The Need for Increased Research on IVR Training and Evaluation

As the interest on IVR training has increased in organizations, recent systematic reviews on IVR training [e.g., 15] point out that it is crucial to understand the relationship between the design of IVR training environments with the training objectives and their underlying pedagogies. Evaluating the technical aspects of IVR technology alone is not sufficient if the learning outcomes of the training scenarios are not meaningful for the users' professional practice [6], when the meaningfulness (e.g., relevancy, usefulness) of their learning outcomes is the key to success [5, 12-13].

Within design science, researchers have paid extra attention to developing evaluation procedures. Seminal models and frameworks [16-17] highlight evaluation as a key activity that is integrated with cycles of design and formalization of learning. However, most of the prior work on DSR-evaluation has focused on the evaluation of traditional technologies [e.g., 18], and thus the DSR evaluation of immersive technologies and the training experiences they provide is an emergent topic for IS [19].

The immersive experience within IVR training environments is very distinct from traditional forms of Information Technology (IT)-mediated experiences of education and training, such as e-learning [20]. The immersive training experience should provide the feeling of becoming completely absorbed and surrounded by an enclosing IVR space, increasing the level of interactivity and sensations of presence among users [21]. Consequently, prior research on VR and education/training [1-4] indicate that evaluation of IVR training needs to define evaluation goals that: (i) address issues that are generalizable on a meta-level (e.g., meta-design, meta-requirements) and impact the design of immersive training environments; (ii) focus on the learning objectives from the perspective of users' first-hand mode immersive experience (e.g., being in the immersive training environment together with the user); and (iii) evaluate the learning performance and outcomes. Knowledge about emerging challenges of organizing and conducting such form of evaluation in the context of IVR and education/training, is however currently understudied [13] and thus needs more attention through further research.

3 A Case of Evaluating IVR Training: Fire Safety Onboard Trains

The study is a part of an ongoing ADR case with the objective of designing and evaluating IVR fire safety training onboard trains. ADR [14] was selected from the beginning of the project as the overarching approach because it provides a sufficient method and framework for organizing and conducting a variant of DSR that emphasizes: (i) processes for building, intervening, and evaluating (BIE) IT-artifacts in organizational settings; and (ii) produce prescriptive knowledge that can be formalized and proposed as design principles. The case is a cooperative effort with the largest train company in Sweden, SJ, and this research in progress reports experiences of a study that was conducted in the on-going project between 2020 and 2021. The case has undergone a full cycle of ADR and produced both research and organization contributions (as shown in Figure 1), whereas this paper emphasizes the evaluation aspect of the project, which is also the current state of the project, namely, to formalize knowledge

from the evaluation phase. This paper is thus reporting an aspect of that knowledge with a particular focus on knowledge about the evaluation challenges that were identified in the study.

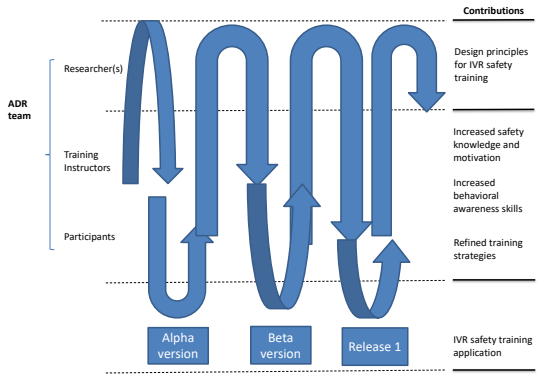

Fig 1 Action Design Research Cycle (adapted from [14])

The goal is to evaluate the learning outcomes of IVR fire safety training for practitioners that work onboard trains (e.g., drivers, conductors). Subsequently, an evaluation plan was framed according to Sein et al's [14] guiding principle 5, 'Authentic and Concurrent Evaluation', which advocates evaluation opportunities that follow natural controls due to the emergent nature of an IT artifact. Hence, in ADR, decisions about designing, shaping, and reshaping the artifact and intervening in organizational work practices, should be "*[…] interwoven with ongoing evaluation, although their specific format may vary based on the BIE form"* [14, p. 43]. We followed this guiding principle and used Venable et al's [16] framework for DSR evaluation, by (i) defining a set of evaluation goals (e.g., *what* needs to be evaluated and *why*), (ii) employing specific techniques for evaluation (e.g., *how* to evaluate the evaluand), and (iii) deciding when to evaluate.

Evaluation Component	**Description**
Evaluation Goals	- *What* to evaluate: safety performance quality, learning experience and outcomes, procedural knowledge development, behavioral skills - *Why to evaluate*: increase the knowledge about deficiencies of design goals, mapping of IVR features with quality of performance, learning outcomes for the participants and organizational value
Evaluation Techniques	- *How* to evaluate: observations, group interviews, web survey
Evaluation Period	- *When* to evaluate: ex-post evaluation through on-site training sessions after the alfa-prototype was built and implemented

Table 1 Overview of the Evaluation Plan: IVR Fire Safety Training

The evaluation plan shown in Table 1 was executed continuously during the ADR cycle together with four distinct groups of involved stakeholders (as shown in Table 2).

Group	**Involved Stakeholders**
Group 2	- 5 participants (2 train drivers + 3 train conductors) - 1 instructor
Group 2	- 6 participants (train drivers) - 1 instructor
Group 3	- 8 participants (3 train drivers + 2 managers + 3 train conductors) - 1 instructor
Group 4	- 5 participants (3 train drivers + 2 train conductors) - 1 instructor

Table 2 Summary of Involved Stakeholders in IVR Training Sessions

The evaluation was executed during the second ADR stage, 'Building, Intervention, and Evaluation' [14, p.41], through group interviews, survey, and on-site observations.

The group interviews were organized and executed according to the following steps. In the **first step**, each group had an instructor that informed the groups about the process of safety training in IVR, showed the groups how to equip IVR devices and provided

the groups a demonstration of the training scenario in IVR. The training scenario was set onboard a virtual train in the IVR environment, with the objective of identifying fire onboard train and extinguishing it as quickly as possible. As a **second step**, each participant of a group performed their safety training scenario. Each performance lasted between 3-8 minutes (Figure 4 shows one of the training participants). As a **third step**, the participants participated in a semi-structured group interview. As a **fourth step**, each interview started with background questions such as: "Have you ever used IVR before?", "What is your prior experience of IVR and training?", "What is your professional role/title?"

In addition to the group interviews, a web survey was sent out to all participants after the training sessions where the participants were asked to answer open questions such as: "What is your overall experience of the IVR training?", "In what ways did you find it useful for your work?", "Would you like to conduct other physical training exercises in IVR as well?". Moreover, on-site observations were made by the principal researcher of this paper to collect data on the IVR training scenario, participants' performance and learning outcomes, and challenges they encountered during the training scenario.

4. Emerging Challenges for DSR Evaluation of IVR Training

Our findings focus on outcomes from the evaluation phase with a particular emphasis on challenges for DSR evaluation of IVR training. A key finding is the importance of *where* the evaluation is conducted because it is qualitatively different to evaluate the performance of IVR training within the immersive setting, than it is to just merely observe the training scenario through a screen, in a physical setting, or online (e.g., e-learning).

The immersive learning experience is holistic and encompasses an increased sense of embodiment and presence, and within an IVR setting this means multimodal features (e.g., text, sound, immersion, avatar guidance, embedded video clips) for interaction and learning [11-13]. Emphasizing the *where* the evaluation needs to take place, supplements thus the *what*, *why*, *how*, and *when* questions which, Venable et al [16] stress as important to address for DSR evaluation. Hence, in this section, we outline three challenges of evaluating IVR training, that we claim will emerge differently depending on *where* the evaluation takes place – e.g., evaluation that takes place within the IVR setting (e.g., through first-hand observation) versus evaluation that takes place in a distinct setting (e.g., through group interviews afterwards) - and what the evaluation goals are – e.g., training performance versus training objectives and the experienced meaningfulness of learning outcomes for participants' profession. The challenges are depicted in Table 3.

Emerging Challenge	Why the Challenge Emerges
1. Verifying Eliciting Motivation and Learning through Immersion in the IVR Environment	The challenge emerges due to the complexity of evaluating the IVR training performance within the boundaries of an IVR environment
2. Validating Acquisition of Cognitive Psychomotor and Affective Skills in the IVR Environment	The challenge emerges due to the difficulty of validating acquisition of cognitive psychomotor and affective skills in the IVR environment
3. Pin-Pointing the Pedagogical Utility and Value of Increased Presence and Embodiment in the IVR Environment	The challenge emerges due to the overshadowing of pedagogical utility and value when immersion occurs through increased presence and embodiment in the IVR environment

Table 3 Emerging Challenges of Evaluating IVR Training

The **first challenge** is pertinent to multimodal features that mediate the game-based training, simulations, and virtual worlds of IVR training, all of which are mediated through immersion to elicit motivation and learning through education/training [13]. Our observations showed us that while the immersive experience provides a high degree of realism, it is difficult to evaluate the motivation behind participants' performance unless the evaluator is situated in the IVR-environment together with the participant, experiencing the implications of immersion. Additionally, due to the versatility of multimodal features in an IVR environment, evaluation of training

performance needs to be mapped with the pedagogical value of the features, and if they contribute or distort the participants' overall training performance. One of the participants problematized the issue of distortion by saying that: *"It [immersion] felt like I was sucked into another world because I could no longer see my surrounding in the physical room. You must be there [the IVR environment] in order to really feel this feeling of being absorbed. The feeling was very strong and it took a lot of attention to get by it for the training"* (Train Driver/Group 2)

The **second challenge** focuses on the effect that the IVR training has towards acquisition of cognitive psychomotor (e.g., the connection between understanding how to move in order to extinguish fire onboard a real train versus doing so in the IVR setting) and affective skills (e.g., the connection between learning outcomes and how meaningful they are for participants' individual knowledge interests) [29]. Our data from group interviews revealed a tension between expectations of learning outcomes (e.g., from the organization's perspective) and actual learning experiences of the participants', making it challenging to evaluate the value of training performance until anomalies of learning outcomes are identified in the real world [7-8]. One of the instructors elaborated on this issue by saying that: *"Safety is not only important for us personnel onboard the train but first and foremost for our passengers. We must take care of them and look after their different needs during fire evacuation. But the current training scenarios do not include other people onboard the train, making it quite simple for us to extinguish the fire in the virtual training"* (Instructor/Group 1

The **third challenge** highlights how easy it is for participants of IVR training to get absorbed by the sensations of increased presence and embodiment. Some of the participants felt ill due to the high degree of presence, making it difficult for participants of an IVR training environment to perform training exercises adequately [11], whereas others were caught up by the fancy interaction opportunities the IVR environment afforded (e.g., teleportation, moving through walls). Our data from observing the training performance of participants via a screen revealed that it is difficult to pin-point what pedagogical utility and value the immersive experience provides for the learning outcomes. The secondary perspective (observing via the screen) omits the first-hand experience of embodiment and increased presence, which makes it difficult to evaluate the learning outcomes from participants' perspective [2]. Data from our group interviews reinforced the validity of this issue when one of the participants revealed that: *"I have never used virtual reality before so for me it felt overwhelming. I had first a difficulty focusing on the training but then after a couple of rounds I started to just focus on extinguishing the fire and not all instructions. I am not sure if I learned anything more than I already knew prior to the exercise, more than how it feels to be in virtual reality"* (Train Conductor/Group 3)

Concluding Remarks

In this research in progress, we have outlined a set of new requirements for design science evaluation of artifacts that are employing IVR technology for education/training purposes. We believe that this kind of evaluation criteria is needed because IVR is markedly different from traditional information systems from the design, use, and experience perspective. Furthermore, the IVR training environment evaluation poses challenges on several levels, as the design of the technical environment is a challenge by itself, and it should not hinder the learning of the often-critical skills that are being learned. To finalize this research in progress, we plan to focus on the following aspects:

- **Evaluation techniques**: more knowledge on specific techniques that are elaborated for evaluation of IVR education/training, is needed. Here, we believe that it is important to test different forms of evaluation to understand which techniques that are congruent with the nature of an IVR experience, and which of them that enable us to evaluate training scenarios together with the participants in IVR. This would allow us to get first-hand experience of testing evaluation techniques that might help us tackle the emerging challenges outlined in Table 3
- **Evaluating learning effectiveness:** determining the quality of learning outcomes and their direct meaningfulness (e.g., relevancy, implications)

for the participants' profession, is needed. Here, we plan to conduct additional observations and interviews that focus the linkage between participants' training performance and the meaningfulness of their learning outcomes (accordingly to the participants' own words and experience)

- **Extracting design principles for IVR training environments that guarantee learning:** extracting design principles that help us address meta-issues such as meta-design and meta-requirements. Here, we plan to formalize findings from the previous point into design principles that prescribe knowledge on how to design meaningful IVR training environments
- **Testing of the above in a real environment**: testing and refining the design principles through additional ADR cycles in future research

References

1. Carruth, D. W. (2017). Virtual reality for education and workforce training. In *15th International Conference on Emerging eLearning Technologies and Applications (ICETA),* 1-6.
2. Feng, Z, González, V. A., Amor, R., Lovreglio, R., & Cabrera-Guerrero, G. (2018). Immersive virtual reality serious games for evacuation training and research: A systematic literature review. *Computers & Education*, *127*, 252–266.
3. Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, *23(4)*, 1515–1529.
4. Anthes, C., Garcia-Hernandez, R.J., Wiedemann, M., Kranzlmuller, D.: State of the art of virtual reality technology. In: 2016 IEEE Aerospace Conference, pp. 1–19 (2016)
5. Fromm, J., Radianti, J., Wehking, C., Stieglitz, S., Majchrzak, T. A., & vom Brocke, J. (2021). More than experience?-On the unique opportunities of virtual reality to afford a holistic experiential learning cycle. *The Internet and Higher Education*, *50*, 100804.
6. Vergara, D., Extremera, J., Rubio, M. P., & Dávila, L. P. (2019). Meaningful learning through virtual reality learning environments: A case study in materials engineering. *Applied Sciences*, *9*(21), 4625.
7. Checa, D., & Bustillo, A. (2020). A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications*, *79*(9), 5501-5527.
8. Çakiroğlu, Ü., & Gökoğlu, S. (2019). Development of fire safety behavioral skills via virtual reality. *Computers & Education*, *133*, 56-68.
9. Howard, M. C., & Gutworth, M. B. (2020). A meta-analysis of virtual reality training programs for social skill development. *Computers & Education*, *144*, 103707.
10. Zhang, H., He, X., & Mitri, H. (2019). Fuzzy comprehensive evaluation of virtual reality mine safetytraining system. Safety Science, 120, 341-351.
11. Joshi, S., Hamilton, M., Warren, R., Faucett, D., Tian, W., Wang, Y., & Ma, J. (2021). Implementing Virtual Reality technology for safety training in the precast/prestressed concrete industry. *Applied ergonomics*, *90*, 103286
12. Fromm, J., Radianti, J., Wehking, C., Stieglitz, S., Majchrzak, T. A., & vom Brocke, J. (2021). More than experience?-On the unique opportunities of virtual reality to afford a holistic experiential learning cycle. *The Internet and Higher Education*, *50*, 100804.
13. Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, 103778.
14. Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS quarterly*, 37-56.
15. Unnikrishnan, R., Konstantinos, K., & Francesco, C. (2021). A systematic review of immersive virtual reality for industrial skills training, Behaviour & Information Technology, 40:12, 1310-1339
16. Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a framework for evaluation in design science research. *European journal of information systems*, *25*(1), 77-89.
17. Peffers, K., Rothenberger, M., Tuunanen, T., & Vaezi, R. (2012, May). Design science research evaluation. In *International Conference on Design Science Research in Information Systems*(pp. 398-410). Springer, Berlin, Heidelberg.
18. Miah, S. J., Hasan, N., Hasan, R., & Gammack, J. (2017). Healthcare support for underserved communities using a mobile social media platform. *Information Systems*, *66*, 1-12.
19. Wohlgenannt, I., Simons, A., & Stieglitz, S. (2020). Virtual reality. *Business & Information Systems Engineering*, *62*(5), 455-461.
20. Attwell, G. (2006). Evaluating E-learning: A Guide to the Evaluation of E-learning. *Evaluate Europe Handbook Series*, *2*(2), 1610-0875.
21. Mütterlein, J. (2018). The three pillars of virtual reality? Investigating the roles of immersion, presence, and interactivity. In *Proceedings of the 51st Hawaii* international *conference on system sciences*.