ASSESSING CHEMICAL INDUSTRY EMPLOYEES USING IMMERSIVE VIRTUAL ENVIRONMENT

Ryo Toyoda

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
School of Engineering
Newcastle University
United Kingdom

January 2023

This project has received funding from the European Union's EU Framework Programme for Research and Innovation Horizon 2020 under Grant Agreement No 812716

Abstract

The highly complex nature of chemical process industry necessitates regular, updated, and relevant training for chemical operators to ensure that they are equipped to perform their functions. Hence, this study proposes the use of immersive virtual reality (IVR) in chemical industries, as they can facilitate effective training in a safe and controlled environment. Following the rapid advancement and growing market of IVR, it becomes necessary to develop an unbiased, and unobtrusive assessment design in this environment. The literature review conducted in this study suggests that VR-based health and safety training, *vis-à-vis* the conventional methods, is more effective in developing and enhancing the reaction, learning, and behavioural levels of trainees. As to the behavioural intention to adopt IVR, data collected and analysed from 438 respondents, who were grouped according to prior IVR experience, nationality, and length of work experience, shows no statistically significant differences among the groups. Since there is lack of development of a systematic assessment framework that is tailored-fit for IVR, this study proposed an assessment structure incorporating and harmonizing appropriate methodologies. The proposed framework uses Dreyfus model, which provides for the progression of a novice to expert, incorporates evidence-centred design (ECD), to provide logistical and evidentiary links from collected data to the desired evaluation construct, and integrates fuzzy comprehensive evaluation method to further enhance the resolution of levels in Dreyfus model. This framework was tested and validated through an IVR-training prototype called 'Operate Your Own Reactor', which was specifically created for this study, in ACTA (Belgium). Lastly, the study discovered that in terms of attitude, satisfaction based on content, and behavioural intention, there are no significant differences whether trainees are presented with simplified or detailed assessment results. Nonetheless, trainees are found to prefer the latter as it helps them improve further.

Keywords: Virtual reality, Chemical industry, Health and safety training, Expertise Development, Assessment, Adoption intention

Acknowledgement

As I step out the comfort and face the real world, I extend my sincerest appreciation and warmest thanks…

To **God Almighty**, who protecting me and giving me strength. Thank You for giving me my family, friends, and the people who have touched my heart. Thank You also for all the guidance that you gave to me. I pray that You continuously guide us.

To **my family**, for your undying support, absolute love, ceaseless care, and all-embracing patience to me. Thank you for praying for me and giving me hope whenever my world seems to fall. Thank you for understanding me for all these years. I believe that saying thank you is not enough for all the sacrifices you made. I am very blessed to have you as my parents and I would never cease to thank God for giving me the best one in the world. Giving you happiness and pride was the main reason I strived this hard.

To **Prof. Jarka Glassey** and **Dr. Fernando Russo Abegão**, my thesis advisers, for accepting me as one of your advisees, for understanding and supporting me, and for the knowledge you have imparted to me. Thank you for your suggestions and guidance whenever I had a hard time with my experiment. Thank you for your encouragement every time I feel that I can't finish my experiment. Your guidance has made me become a better person. Thank you for being the best thesis adviser.

To **European Training Network for Chemical Engineering Immersive Training Programme (ETN-CHARMING)**, thank you for the opportunity to be part of this programme. Thank you for the encouragement to achieve excellence not only in academics but also in every aspect in life. CHARMING completes my PhD life with the best academics and social life that serve as my armours as I graduate in this programme. Thank you everyone for being there. I will never forget all of you.

To my fellow CHARMING work package 3 'Operate Your Own Reactor' team members, **Ms. Sofia Garcia Fracaro, Mr. Timothy Gallagher,** and **Ms. Yusra Tehreem**, thank you for your time and your effort to build and develop the OYOR prototype together. In addition, thank you all for your help during the data collection in MERCK and ACTA. Specifically, thank you **Sofia** for creating and validating the overall content. Thank you **Tim** for exerting your effort to develop the appropriate learning analytics to our OYOR prototype. Lastly to **Yusra** for coding and building our OYOR prototype using Unity software.

To **Dr Bert Slof**, thank you for providing feedback on the modified UTAUT2 questionnaire (Appendix A). Likewise, thank you to **Mr. Michael Dumin**, **Ms. Michaela**

Arztmann, **Mr. Philippe Chan,** and **colleague of Mr. Chan at ARKEMA** for translating the questionnaire.

To **Ms. Sofia Garcia Fracaro** and **Dr. Michael Wilk** of **Merck KGaA** and **Mr. Philippe Chan** and **Dr. Jean-Luc Dubois** of **ARKEMA**, thank you for your support throughout the UTAUT data collection period.

To **Dr Liesbeth Kester** of Utrecht University (Netherlands), thank you for providing feedback on the IVR-based assessment system preference questionnaire (Appendix D). Moreover, a special thanks goes to **MERCK KGaA health and safety officers** for the time and resources they provided and to all the **MERCK KGaA and ACTA participants** that took the time to test our prototype and answer my questionnaire.

To **Rhea**, thank you for understanding me and unconditional patience and love during my thesis writing days. I am really thankful to have you as my fiancée. Everything was less stressful and a lot happier with you virtually/physically around.

To **CHARMING ESRs**, for the unforgettable bonding time here in Europe. Thank you for winching up my PhD life. Thank you for the consummate friendship and bonding time. Thank you for the coolest adventurous trips. Thank you for the happy hours and non-stop drinking sessions. I will certainly never forget and miss these moments.

Papers presented and published

The following papers were published in the past three academic years (2019-2022).

1. **Toyoda, R.**, Russo-Abegão, F., & Glassey, J. (2022). VR-based health and safety training in various high-risk engineering industries: a literature review. International Journal of Educational Technology in Higher Education, 19(1), 42. https://doi.org/10.1186/s41239-022-00349-3
2. Garcia Fracaro, S., Tehreem, Y., Gallagher, T., **Toyoda, R.**, Glassey, J., Bernaerts, K., & Wilk, M. (2022). Procedure and emergency virtual reality training in the chemical industry: Study of effectiveness. Chemie Ingenieur Technik, 94(9), 1375–1376. https://doi.org/10.1002/cite.202255202
3. Udeozor, C., **Toyoda, R.**, Russo Abegão, F., Glassey, J. (2022) Digital Games in Engineering Education: Systematic Review and Future Trends. European Journal of Engineering Education. 1-19. https://doi.org/10.1080/03043797.2022.2093168
4. **Toyoda, R.**, Russo Abegão, F., Gill, S. & Glassey, J. (2021). Drivers of immersive virtual reality adoption intention: a multi-group analysis in chemical industry settings. Virtual Reality. https://doi.org/10.1007/s10055-021-00586-3
5. Udeozor, C., **Toyoda, R.**, Russo Abegão, F., & Glassey, J. (2021). Perceptions of the use of virtual reality games for chemical engineering education and professional training. Higher Education Pedagogies, 6(1), 175–194. https://doi.org/10.1080/23752696.2021.1951615
6. Garcia Fracaro, S., Chan, P., Gallagher, T., Tehreem, Y., **Toyoda, R.**, Bernaerts, K., Glassey, J., Pfeiffer, T., Slof, B., Wachsmuth, S., & Wilk, M. (2021). Towards design guidelines for virtual reality training for the chemical industry. Education for Chemical Engineers, 36, 12–23. https://doi.org/10.1016/j.ece.2021.01.014

The following papers were presented in the past three academic years (2019-2022).

1. **Toyoda, R.**, Garcia-Fracaro, S., Gallagher, T., Tehreem, Y., Russo Abegão, F., & Glassey, J. (2022) Learning analytics in VR-based chemical training: operator apprentice perspectives. In: CHISA - 26th International Congress of Chemical and Process Engineering, Prague, Czech Republic.

2. Tehreem, Y., Fracaro, S. G., Gallagher, T., **Toyoda, R.**, Bernaerts, K., Glassey, J., Abegao, F. R., Wachsmuth, S., Wilk, M., & Pfeiffer, T. (2022). May I Remain Seated: A Pilot Study on the Impact of Reducing Room-Scale Trainings to Seated Conditions for Long Procedural Virtual Reality Trainings. In E. Wu, Z. Pan, & A. Song (Eds.), 2022 8th International Conference on Virtual Reality (ICVR) (pp. 62–71). IEEE. https://doi.org/10.1109/ICVR55215.2022.9848222
3. Chan, P., Guilpain, G., Dubois, J., **Toyoda, R.**, Russo Abegão, F., Glassey, J., Gallagher, T., Kester, L., Bernaerts, K., van Gerven, T. (2021) VR LaboSafe Game: A Virtual Reality Training Game for Chemical Engineering Laboratory Safety. In: 13th European Congress of Chemical Engineering and 6th European Congress of Applied Biotechnology. Virtual: IEEE.
4. van Gerven, T., Dumin, M., Dominguez Alfaro, J., Solmaz, S., Santos, P., Smets, I., Bernaerts, K., van Puyvelde, P., Nasser, R., van Loenen, S., Adriaensen, K., Haelterman, K., Arztmann, M., Hu, Y., Gallagher, T., Hornstra, L., Slof, B., Wouters, P., Kester, L., Croitoru, N., Fornos, S., Cermak-Sassenrath, D., Udeozor, C., **Toyoda, R.**, Russo Abegão, F., Glassey, J., Chan, P., Guilpain, G., Dubois, J., Garcia Fracaro, S., Wilk, M., Tehreem, Y., Pfeiffer, T. (2021) Demonstration of Immersive Tools for Chemical Engineering Training. In: 13th European Congress of Chemical Engineering and 6th European Congress of Applied Biotechnology. Virtual Event: ECCE-ECAB.
5. **Toyoda, R.**, Russo Abegao, F., Gill, S., Glassey, J. (2021) A Cross-Cultural Study on the Perception of the Use of Immersive Virtual Reality in Health and Safety Training for Professional Chemical Engineers. In: ChemEngDay 2020. Bradford, UK.
6. **Toyoda, R.**, Russo Abegão, F., Gill, S., & Glassey, J. (2020). Perceptions of Professional Chemical Engineers toward Immersive Virtual Reality in Health and Safety Training. The European Conference on Education 2020: Official Conference Proceedings, 273–284. https://papers.iafor.org/submission57993/

Furthermore, the following papers will be submitted to journals:

1. A paper entitled "A framework for evaluating user expertise in immersive learning environments" which provides an instruction on employing an assessment framework for the evaluation of users in IVR learning environment (Chapter 4). This paper has been submitted to the Assessment in Education: Principles, Policy & Practice Journal (September 2022).

2. A paper entitled "Incorporating assessment into chemical industry training through immersive virtual reality (IVR): a case study" which determines the overall preference of a user as to the details presented in the assessment results (Chapter 5). This paper has been submitted to the Education and Information Technologies Journal (September 2022).

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Background and significance of the study

High-risk industries such as chemical processes, are usually highly complex and contain numerous interacting variables such that changing one may affect several variables simultaneously (Colombo *et al.*, 2014; Garcia Fracaro *et al.*, 2021). Due to these complexities, chemical plant operators face additional challenges since erroneous interpretations or assumptions made during the operations may lead to a serious accident (Nazir *et al.*, 2012; Patle *et al.*, 2019). Since handling and management of accidents do not always proceed exactly as written and described in manuals, it is essential for the operators to have a good level of awareness and the ability to perform accurate actions to ensure productivity and safety of the plant both in normal and abnormal situations (Colombo *et al.*, 2014; Zare *et al.*, 2019).

To ensure the safe operation of each equipment in a chemical plant, operators must be able to interpret the available information, collaborate with different teams, understand the risks and consequences of every action, and deal with the given situation based on their experience (Colombo & Golzio, 2016). For the operators to acquire these skills, they must undergo a series of effective training, such as health and safety (H&S) training. In the past, a combination of various training methods, such as classroom training, on-the-job training, and/or 2D simulation tools, were considered to be an effective training method to enhance the required skills for operators and employees (Patle *et al.*, 2014, 2019).

However, for highly automated chemical plants these training methods alone are insufficient in effectively training the operators as lecture-based training has drawbacks with poor engagement of learners, on-the-job training may be unsafe, and 2D simulations may not provide enough realism of the process (Patle *et al.*, 2019). In applying these methods, operators may find it difficult to respond adequately and appropriately during unexpected situations in real life. A factor that may contribute to the ineffectiveness of these training methods is their concentration on troubleshooting the emergency situation rather than understanding the rationale of why an operation or process is risky in the first place (Kumar *et al.*, 2021; Thompson & Falco, 2020).

Given the abovementioned challenges associated with conventional training methods, the use of technology-aided learning materials, such as virtual reality (VR) technology is receiving great interest from different stakeholders. The recent advancement in the field of VR technology enables the creation of more realistic, in-depth, safe, yet complex 3D chemical plant scenarios, both normal and/or dangerous conditions, where trainees can enhance their anticipation of circumstances through participation (Manca *et al.*, 2013; Shamsuzzoha *et al.*, 2019). For instance, aside from the ability to create dynamic, immersive, and 3D simulated chemical plant environments where professionals can interact and move freely, VR technologies allow users to have a better understanding of the schematics of the process/plant which are abstract at the representation level (Feise & Schaer, 2021; Nazir *et al.*, 2012). Although research focusing on the application of VR in terms of feasibility, usability, *etc.*, has received considerable attention, only limited analysis of different assessment methods and instruments to evaluate the effectiveness of VR-based training in developing skilled operators is currently available.

In order for the stakeholders, especially trainers and administrators, to make sure that a specific VR-based training is effective and aligned with the needs of their organisation it is vital for them to understand the various assessment methods used to evaluate their effectiveness. For instance, the advances in VR-based technologies, such as desktop computers and head-mounted displays (VR-HMDs), have made it possible to trace and capture several in-game learner-generated data that can be used as indicators to assess their corresponding knowledge and/or skills (Drey *et al.*, 2020; Min *et al.*, 2019; Shute & Rahimi, 2021). Through examining questions like why and/or how learning takes place when using VR-based training, it is possible to measure the degree to which the chosen VR-based training equips the trainees in accomplishing certain tasks as well as addressing unforeseen circumstances.

It is also imperative to verify the adoptability of the new VR-based training in the perspective of the operators and employees working in chemical industries. According to the critical mass theory proposed by Oliver *et al.* (1985), sufficient numbers of individuals (*i.e.* the critical mass) who are willing to try and use readily available new technology is necessary in order to achieve a self-sustaining rate of adoption and to create network-like benefits (Lew *et al.*, 2020). Through uncovering the demand of the users who are keen to adopt VR-based training, stakeholders will have a clearer view on what specific construct(s) (*e.g.* more feedback, more game elements, more easy controls, *etc.*) should be prioritised in designing appropriate training in the VR environment.

Given that the VR-based technologies have the potential to play a significant role in delivering professional development, several stakeholders have begun asking for the development of useful training metrics which will utilise user-generated data that may be used to assess whether the specified learning outcome(s) were achieved (Loh *et al.*, 2015). Currently the guidance on good practices in terms of employing assessment in VR-based environment is very limited. This is due to the complex nature and novelty of the environment requiring a methodological and tedious process of creating an appropriate assessment technique. Thus, developing a simple and easy to follow guidelines on improving the design and utility of these digital assessment systems would be highly beneficial to the community.

In order to capitalise on the potential of the VR-based training in terms of its assessment, stakeholders should consider factors such as (1) how to extract, organise, and quantify the most important information from the pool of data in the VR environment, (2) how to categorise the degree of learning shown by the trainees in the VR environment, (3) how to verify that the collected information from the VR environment provides sufficient and clear interpretations related to the learning and development of trainees, (4) how to calculate and present the performance results as well as the corresponding feedback of the trainees after the VR-based training.

Given that the implementation of every important construct requires substantial amount of time and effort, it will be helpful for the stakeholders to follow an effective, accessible, and reliable set of guidelines to speed up the implementation of assessment systems in the VR-based training. A case-study aiming to understand how a set of developed and validated assessment guidelines can be integrated in the VR-based training by different stakeholders for their respective accountability purposes (*e.g.* evaluation of trainees, identification of skilled workers) would be beneficial in achieving this goal.

1.2. Aims and objectives

The main objective of this study is the establishment of an assessment framework that can be integrated in VR-based training environments to assess the skills of chemical industry operators. To achieve this, there are several objectives that need to be completed:

1. To conduct a systematic examination of existing literature on the current studies in VR-based health and safety training, assessment techniques, training evaluation, and the potential of these VR-based technologies to improve the training evaluation outcomes in various high-risk industries.
2. To explore the similarities and differences between the different sub-categories of chemical operators and employees in terms of their intentions to adopt VR-based training.
3. To develop a guideline on employing an unbiased and unobtrusive assessment framework for the evaluation of user expertise in immersive learning environments.
4. To perform a case-study with industrial partners (MERCK and ACTA) in order to identify the potential benefits of using the developed assessment framework in VR-based chemical process safety training.

5. To determine the preferences of the users regarding how digital assessment should be implemented in VR-based chemical process safety training.

1.3. Thesis outline

This thesis contains details of research undertaken in partial fulfilment of the Doctor of Philosophy (PhD) award in Chemical Engineering from Newcastle University. All experimental work was led by the primary author, Ryo Toyoda. The list of individuals who helped the primary author during the experimental work can be found in the Acknowledgement section. All of the chapters were written by the primary author, and edited by Dr. Fernando Russo Abegão and Prof. Jarka Glassey. It is divided into four major chapters in addition to the introduction and conclusions.

This introduction is followed by a systematic literature review (Chapter 2) which focuses on exploring the effectiveness of VR-based training. The review includes a brief discussion of the history, advantages as well as challenges of VR. Current application of VR-based training in various high-risk industries is also reviewed. In addition, discussions on the different assessment methods as well as their application in different VR-based training in various high-risk industries are presented. Furthermore, a review on the potential of VR-based training to improve the training evaluation outcome(s) compared to traditional and/or other VR-based training methods is provided. Lastly, this review chapter proposes a training design framework that may be adopted to align the VR-based training with desired training outcome and assessment method.

Chapter 3 investigates the interrelationship between influential factors and behavioural intention to adopt immersive virtual reality (IVR) among different groups of people working in chemical industries through adapting the modified Unified Theory of Acceptance & Use of Technology 2 (UTAUT 2). Data were collected using an online questionnaire from chemical operators and/or employees. Partial least squares structural equation modelling (PLS-SEM), based on SmartPLS^{TM} version 3, was used to analyse the intention of the population sampled to adopt IVR, followed by multi-group analysis (MGA) to explore the differences between groups of chemical operators and/or employees. The findings from this chapter contribute to the literature of the UTAUT2 model on IVR adoption in formulating appropriate strategies to improve the adoption of IVR in different group settings.

Chapter 4 discusses the detailed procedure of establishing an effective assessment framework for IVR-based training comprised of a three-step approach. The first step involves the creation of competency-based assessment rubrics in IVR setting through employing a combination of methodologies such as the evidenced-centred design (ECD) model, and the Dreyfus model. Afterwards, the developed assessment framework is validated by a group of experts in the H&S training through a focus group discussion (second step). Subsequently, the overall user performance score is evaluated using a fuzzy comprehensive evaluation method (third step). The developed assessment framework is then implemented and tested in a case study which involves an IVR-based chemical process safety training called ‘Operate Your Own Reactor’. The results from this chapter confirmed that the developed assessment framework, which was integrated in the IVR-based safety training system on the production of *n*-butyllithium, has positive implications for various stakeholders such as trainees, trainers, and administrators.

Since the success of implementing digital-based assessment in IVR environment depends on the delivery and communication of assessment results to the learners, as well as their corresponding feedback, Chapter 5 aims to evaluate the perspective of the participants with regard to the delivery and presentation of their performance in IVR-based training. In pursuit of the abovementioned objective, a quasi-experimental approach was implemented. The participants were divided into a control group and an experimental group. The control group performed the IVR-based training showing simple assessment system (*e.g.* only overall grade) while the experimental group performed the same IVR-based training showing detailed assessment system (*e.g.* more information such as number of mistakes, or hints per sub-task aside from the overall grade). The findings from this chapter provide insights for assessment designers on how evaluation should be presented in IVR environment in order to support autonomous learning.

The overall impact of the work described in the previous chapters is discussed and summarised in Chapter 6. The recommendations for future work are also discussed in this chapter.

Chapter 2. VR-based health and safety training in various high-risk industries: a literature review

The rapid evolution of the digitalised world is contributing to the rise in reports on the benefits and applications of virtual reality (VR) for health and safety training in various high-risk industries. However, the degree of effectiveness of these VR training methodologies has yet to be fully explored. To bridge this gap, a critical review of the existing VR-based studies was conducted. Unlike previous review studies that focused on uncovering the characteristics of the technology and the challenges in this field, this study mainly evaluates the effectiveness of VR-based training according to Kirkpatrick's model (Kirkpatrick, 2006) as it is the most commonly adopted framework for helping stakeholders to analyse their training programmes. A total of 1516 records were identified through an explorative search, conducted in Scopus, of relevant citations from 2011-2021. 59 articles were considered and classified based on (1) the topics of H&S training, (2) the type of VR, (3) the types of training evaluation outcome, (4) the types of assessment techniques and (5) its potential to improve the training evaluation outcomes. The results of this analysis indicate the breadth of VR-based applications in training users on a combination of topics including risk assessment, machinery and/or equipment process/procedural operation or both topics in various industries. Data showed that the use of fully immersive VR increased significantly due to the improvements in hardware, display resolution, and affordability. Most of the articles used external assessment to measure the changes in the satisfaction and the declarative knowledge of trainees as these are easier to implement, while some articles started to implement internal assessment that provides an automated assessment capable of measuring complex skills. The results of the study also suggest that the VR-based H&S training has the potential to improve the training evaluation outcomes compared to traditional training methods. The findings from this study provide practitioners and safety managers a training design framework that may be adopted to optimise the condition of a VR-based training.

2.1. Introduction

High-risk industries operating potentially dangerous processes, such as those found in the aviation, construction, and chemical sectors, must comply with occupational health and safety standards to sustain nearly error-free levels of performance (Health and Safety Executive., 1999; Hudson, 2003). For an organisation to maintain its high standards, it is important to create and provide a well-developed experiential health and safety (H&S) training programme. However, for the abovementioned industries, it is often impractical to further supplement experiential learning using traditional approaches due to reasons such as cost, operation, and practical limitations (Gao *et al.*, 2019).

Given these challenges associated with the traditional H&S training methods, the importance of technology-aided learning material has been increasing (Soret *et al.*, 2019). Among the different technology-aided learning materials, virtual reality (VR) is becoming more widely integrated into H&S training for safety-critical industries. This technology can create safe yet complex learning and training environment, as well as promote knowledge acquisition through active involvement (Gao *et al.*, 2017; Isleyen & Duzgun, 2019). Considering the promising learning improvement that VR can provide, it is understandable that the number of publications focusing on the application of VR to specific areas such as education and training is increasing (Checa & Bustillo, 2019; L. Jensen & Konradsen, 2018). However, little research has been conducted into the analysis of different assessment methods and instruments to evaluate the effectiveness of VR training.

To make sure that the new VR training is effective and aligned with the needs of the organisation it is important for the stakeholders, especially trainers and administrators, to understand why and how learning occurs when using VR tools in order to choose appropriate assessment methods and use these as indicators of the effectiveness of the VR training. However, the recent review on the effectiveness of both conventional and computer-aided technologies for health and safety training in the construction sector by Gao *et al.* (2019) stated that the empirical evidence supporting the effectiveness of computer-aided technologies (CAT) is still limited. This claim was based on the result that out of the 34 CAT articles considered, only one study evaluated the effectiveness of knowledge acquisition during training (Gao *et al.*, 2019).

Professional training is considered effective when the required attributes, such as problem-solving and analytical skills, are transferred and applied successfully to the daily jobs of trainees. Since it is the responsibility of stakeholders to choose and implement effective health and safety training, it is beneficial for them to understand the various assessment methods used to evaluate training effectiveness. The purpose of this study is to conduct a systematic examination of literature to review the assessment methods used to evaluate the outcomes of the different VR-based health and safety programmes in various high-risk industries. Thus, the following research questions guided this review:

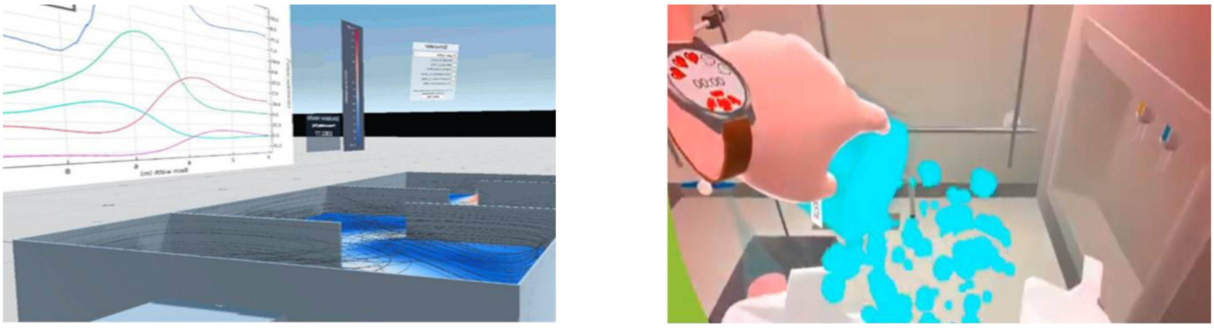
(1) What topics have researchers investigated for VR-based health and safety training in various high-risk industries?

(2) What types of VR were used to deliver health and safety training in various high-risk industries?
(3) What were the outcome(s) measured for establishing the effectiveness of the VR-based health and safety training in various high-risk industries?
(4) What assessment techniques were used to evaluate the outcome(s) of VR-based health and safety training in various high-risk industries?
(5) Does VR-based health and safety training in various high-risk industries have the potential to improve the training evaluation outcome(s) compared to traditional and/or other VR-based training methods?

2.2. Virtual reality

Although VR environment is simulated artificially, this tool has been progressively used for different training purposes as it allows trainees to behave and react as close as possible to reality where they can construct their knowledge through observing, manipulating, and analysing information obtained inside a safe and interactive environment (Avveduto *et al.*, 2017; Kassem *et al.*, 2017). According to Burdea (2003, p3), the term *"virtual reality (VR) is a high-end user-computer interface that involves real-time simulation and interactions through multiple sensorial channels"*.

VR applications which are used for educating and training professionals are usually presented either as games with a serious purpose (*i.e.* serious games) or as simulations (Imlig-Iten & Petko, 2018; Menin *et al.*, 2018). Serious game is a type of screen-based video game with primary purpose of educating about a specific field and supply substantial learning outcome to its users, rather than those which are designed for leisure and/or recreation (Feng *et al.*, 2018). Meanwhile, simulation is a type of a virtual scene application which allows users to change various variables and experience the recreated representation/replica of the system for the purpose of knowledge acquisition and skills development (Imlig-Iten & Petko, 2018). Examples of VR-based simulation and serious game are shown in Figure 2.1.



(a) (b)

Figure 2.1. Example of VR-based (a) simulation and (b) serious game.

As Imlig-Iten and Petko (2018) mentioned, some similar features between serious games and simulations include 3D models of real-life situations and educational frameworks content, interactive and contextual experience exercises that allow users to apply real world knowledge, skills, and strategies. However, there are still fundamental differences between these two technologies as shown in Table 2.1 (Imlig-Iten & Petko, 2018).

Table 2.1. Comparison of Simulations and Serious Games (Imlig-Iten & Petko, 2018).

Simulations	**Serious Games**
Composed of simple rules and/or controls	Supplemented by simple and complex rules and/or controls
Do not contain any additional features (*e.g.* game-element)	Contain many additional features that generate creative stimulation.
Without competitive factors	With competitive factors (*e.g.* obtaining better scores).

The concept of VR is not a recent technology. In fact, it was first materialized in 1962, when cinematographer Morton Heilig patented his first VR device invention called *Sensorama Simulator* (Burdea, 2003). In this motorcycle simulation, aside from motion, colour, and sound during the ride, the rider could sense the odour of the city, the wind effect, and a seat vibrating when experiencing rough road condition (Burdea, 2003). In 1960, Heilig also patented the first head-mounted display (HMD) without motion-tracking, which provided stereoscopic 3D images with wide vision and stereo sounds which replace the usual cinematographic experience (Burdea, 2003; Heilig, 1962).

In the same period, Ivan Sutherland developed and presented his vision of the concept of creating a highly realistic simulation for its users through HMD that is not limited to sound, smell, and haptic feedback, but that also includes object interaction, known as Ultimate Display (Cipresso *et al.*, 2018; Sutherland, 1965). He also continued Heilig's initial work on HMD by using two cathode ray tubes (CRTs) mounted along the ears of the user and he realized that he could use digital scenes instead of analogue photos (Burdea, 2003). This led to his creation of The Sword of Damocles which used HMD connected to computer where the head position and orientation of the user was tracked in order to update the virtual image (Cipresso *et al.*, 2018). Figure 2.2 shows the chronology of virtual reality development (Garcia Fracaro *et al.*, 2021; Molnár, 2017).

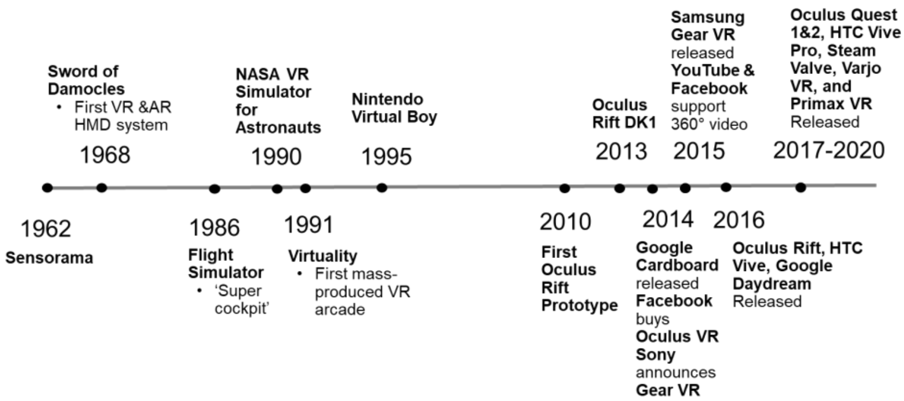


Figure 2.2. Milestones in the history of virtual reality (Garcia Fracaro *et al.*, 2021; Molnár, 2017).

VR technology provides users with a safe 3D training/educational-related space where they can construct their knowledge that reflects real-life situations and events (Fällman *et al.*, 1999). For instance, it is difficult to completely mimic and/or realistically simulate highly dangerous situations such as a rescue operation or a fire incident. However, it is possible to create a representation of real-life scenario using VR technology, which allows firefighters, rescue operators, and the public to be exposed and to be trained in dealing with chaotic catastrophe (Ahmad *et al.*, 2019; Czarnek *et al.*, 2019; Shi *et al.*, 2019). VR technology also provides an environment where its users may gather relevant information and different perspectives through their active engagement from the different virtual scenes (Fällman *et al.*, 1999).

VR technology may also record, monitor, and process the behavioural responses of its users through their body motions and non-verbal expressions while they are immersed in the platform (Alcañiz *et al.*, 2018; Shi *et al.*, 2019). Moreover, several psycho-physiological signals (*e.g.* heart rate, eye movement, *etc.*) can be collected to provide valuable information about cognitive and emotion states (Alcañiz *et al.*, 2018; Czarnek *et al.*, 2019; Soret *et al.*, 2019).

Nonetheless, VR technology, as applied in training, has some notable drawbacks. Since the trainees are immersed in the virtual reality platform, they have reduced awareness in their actual and physical environment. Such reduced environmental awareness may lead to concerns about their safety, security and well-being during training. For instance, some studies suggest that VR users experience motion sickness risking their well-being (Davis *et al.*, 2014; Somrak *et al.*, 2019). To mitigate these potential hazards, the conduct of training must be supervised at all times and the physical space where it is conducted must be wide and unobstructed enough to be safe for the trainees. These mitigating measures require additional costs and resources which are additional to the cost of the technology itself.

In addition, simulation of actual scenarios in a chemical plant may be difficult to create realistically in terms of sensory experience. Chemical plant operators often consider inputs from sensory experience other than sight and hearing, such as heat, weight of materials and equipment, flow of liquid, pressure, among others. These considerations may be difficult to mimic in consumer virtual reality resulting in limited learning in some specific instances. On the other hand, virtual reality may also provide opportunities for the trainees to experience what is impossible to see in reality, such as what is actually happening inside a chemical reactor, dispersion of radioactivity, chemical reactions in atomic level, among others. Table 2.2 lists the key the advantages and disadvantages of VR (Kavanagh *et al.*, 2017).

Table 2.2. Advantages and disadvantages of VR (Kavanagh *et al.*, 2017).

Advantages	Disadvantages
Safe space • Simulates risky scenarios	**Input hardware usability** • Slightly impairs motion and senses
Deeper Learning • Provides relevant knowledge and wider perspective to the user	**Insufficient Realism** • Impedes learning experience if the overall design is unrealistic
Immediate Feedback • Readily imparts the performance evaluation to the user	**Costs** • Requires substantial amount for purchase, maintenance, and support
Increased user motivation and enjoyment • Engages the user better through game-based scenarios	**Training** • Necessitates extensive training for primary users such as trainers and trainees
Social skills • Allows collaboration with other users in various situations	**Lack of Engagement and ineffective** • Discourages interaction if the virtual scenes were poorly designed
Cognitive and constructive skills • Users acquires and retains skills through actual experience	
Personalized learning • Users can digest thing at their own speed	

Kavanagh *et al.* (2017) suggested possible actions to address the limitations of VR technology. For instance, they suggested that the improvement of the ease-of-use of the VR devices may help the users in adjusting more quickly to the training environment. If properly implemented, the user may be able to focus more on the training to maximise learning. Moreover, to enhance realism of the virtual environment, researchers suggested improving latency and graphics. However, this will require more advanced technology and additional cost to implement. Thus, costs and better realism must be deliberately balanced in order to improve the overall usability of VR systems (Kavanagh *et al.*, 2017).

2.3. Methodology

2.3.1. Selection criteria

A detailed review of research studies published within an 11-year time frame (Jan. 2011 to Dec. 2021) was undertaken following the procedure proposed by Kitchenham (2004). The date criterion was based on the need to provide an updated picture of the recent VR development in various high-risk health and safety programmes. A literature search was conducted on Scopus (www.scopus.com) as this is the largest abstract and citation database of peer-reviewed research literature (Jin *et al.*, 2019). The following keywords were used for the literature search:

"Immersive virtual environment", or "virtual reality" or "virtual environment" or "VR", and "assessment" or "performance assessment" or "evaluation" and "health and safety training" or "safety training" or "industrial safety" or "plant operators" or "high risk industry"

The literature search and publication selection processes are shown in Figure 2.3. As shown, the literature search yielded 1516 records. These records were then screened based on the language, year of publication, document type, and whether the reported application was for health and safety training in high-risk industries. After examining the title and abstract of each publication, 178 articles were identified as eligible. These selected publications were then subjected to further full-text screening. Fifty-nine (59) articles were selected for a detailed analysis following the screening criteria such as whether or not the considered article applied a specific assessment method and not merely describing the framework of the projects.

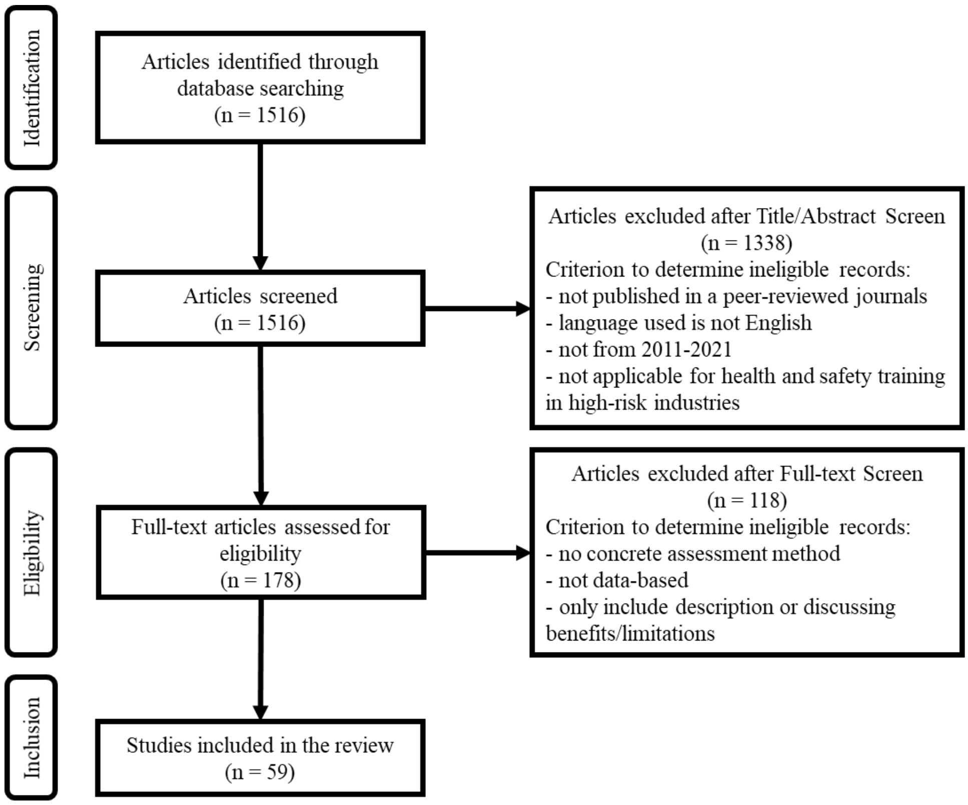


Figure 2.3. Literature search and publication selection process.

2.3.2. Data analysis

In line with the prompts introduced by the research questions, the following information serving as the column headings was extracted from each article (shown in Table 2.3). Research topic pertains to risk assessment (RA) and machinery and/or equipment process/procedural operation (MPO). Level of training evaluation pertains to reaction level (Level 1), learning level (Level 2), behavioural level (Level 3), and results level (Level 4).

Table 2.3. Summary table of each article reviewed.

Authors, Year	**Area of application**	**Research topic**	**Participants**	**Training environment**	**Data Source**	**Assessment category**	**Implementation phase**	**Level of training evaluation**
(Lin *et al.*, 2011)	Construction-related	RA	5 university students	Non-immersive VR	Questionnaire	External	After training	Level 1
(Blickensderfer *et al.*, 2012)	Aviation-related	Both RA and MPO	32 pilots	Semi-immersive VR	Practical test	External	During training	Level 2
					Knowledge test	External	After training	
					Questionnaire	External	After training	Level 1
(Guo *et al.*, 2012)	Construction-related	MPO	15 trainees	Non-immersive VR	Interview + Questionnaire	External	After training	Level 1
(H. Li *et al.*, 2012)	Construction-related	MPO	30 operators	Non-immersive VR	Multiple choice	External	After training	Level 2
					Interview			Level 1
(A. M. D. Jensen *et al.*, 2012)	Clinical or dental-related	Both RA and MPO	15 students and staff members	Non-immersive VR	Questionnaire, interview	External	After training	Level 1
					Think aloud	External	During training	
(Sacks *et al.*, 2013)	Construction-related	Both RA and MPO	66 university students and professionals	Semi-immersive VR	Knowledge test	External	After training and 1 month after training	Level 2
					Questionnaire		After training	Level 1

(Beyer-Berjot *et al.*, 2014)	Clinical or dental-related	Both RA and MPO	10 novice surgeons and 10 experienced surgeons	Non-immersive VR	Log data	Internal	During training	Level 2
(Perlman *et al.*, 2014)	Construction-related	RA	61 university students and professionals	Semi-immersive VR	Knowledge Test	External	After training	Level 2
(Albert *et al.*, 2014)	Construction-related	RA	3 groups of professional engineers (5-12 each group)	Non-immersive VR	Log data	Internal	During training (Compilation of data from 16 work period)	Level 3
					Knowledge test	External	Before training and after 16th work period	Level 2
(Tian *et al.*, 2015)	Aviation-related	MPO	40 students	Non-immersive VR and Semi-immersive VR	Log data	Internal	During training	Level 2
					Questionnaire	External	Before and After training	Level 1
(Nazir *et al.*, 2015)	Chemical Process / Laboratory related	RA	24 university students	Semi-immersive VR	Log data	Internal	During training	Level 2
(Le *et al.*, 2015)	Construction-related	Both RA and MPO	15 participants	Non-immersive VR	Questionnaire	External	After training	Level 1
(Ayala García *et al.*, 2016)	Electric Power and electronics-related	MPO	24 apprentices	Non-immersive VR	Practical Exam	External	9 days after training	Level 3
					Knowledge Test		3 days after training	Level 2
(Ojados Gonzalez *et al.*, 2017)	Agricultural related	MPO	127 participants	Fully immersive VR	Log data	Internal	During training	Level 2
					Questionnaire	External	After training	Level 1
(Gallegos-Nieto *et al.*, 2017)	Manufacturing and assembly related	MPO	15 university students	Non-immersive VR	Log data	Internal	During training	Level 2
(Tawadrous *et al.*, 2017)	Chemical Process / Laboratory related	RA	29 university students	Non-immersive VR	Log data	Internal	During training	Level 2
(Dorozhkin *et al.*, 2017)	Clinical or dental-related	RA	49 professionals	Fully immersive VR	Questionnaire	External	After training	Level 1
(Dado *et al.*, 2018)	Manufacturing and assembly related	MPO	22 university students	Fully immersive VR	Knowledge Test	External	After training	Level 2
(Lin *et al.*, 2018)	Construction-related	RA	43 workers	Non-immersive VR	Knowledge Test	External	After training	Level 2
					Questionnaire			Level 1
(Ogbuanya & Onele, 2018)	Electric Power and electronics-related	MPO	142 university students	Non-immersive VR	Knowledge Test	External	Before and After training	Level 2
(Hjelmervik *et al.*, 2018)	Maritime-related	MPO	17 students	Semi-immersive VR	Log data	Internal	During training	Level 2
(Makransky *et al.*, 2019)	Chemical Process / Laboratory related	Both RA and MPO	105 university students	Fully immersive VR and Non-immersive VR	Practical Exam	External	10 days after training	Level 3
					Multiple choice		After training	Level 2
					Questionnaire		Before and After training	Level 1
(Isleyen & Duzgun, 2019)	Mining Related	RA	5 university staff	Fully immersive VR	Interview	External	After training	Level 1
(Leder *et al.*, 2019)	Manufacturing and assembly related	Both RA and MPO	68 apprentices	Semi-immersive VR	Practical Exam	External	After training and after 6 months	Level 3
					Multiple Choice		After training and after 6 months	Level 2
					Questionnaire		After training	Level 1
(Nicoletti & Padovano, 2019)	Chemical Process / Laboratory related	RA	28 operators	Non-immersive VR	Log data	Internal	During training	Level 2
(Vahdatikhaki *et al.*, 2019)	Construction-related	MPO	5 training experts	Fully immersive VR	Questionnaire	External	After training	Level 1
(Liang *et al.*, 2019)	Mining Related	Both RA and MPO	20 volunteers for part 1 and 10 volunteers for part 2	Fully immersive VR	Log data	Internal	During training	Level 2
(Polivka *et al.*, 2019)	Clinical or dental-related	RA	74 home health care professionals and students	Non-immersive VR	Knowledge test	External	After training	Level 2
					Questionnaire	External	After training	Level 1
(Pham *et al.*, 2019)	Construction-related	RA	40 university students	Non-immersive VR	Knowledge Test	External	Before and after training	Level 2
			60 participants		Interview + Questionnaire		After training	Level 1

			(10 educator, 10 construction professionals , 40 students)					
(Choi *et al.*, 2020)	Construction-related	MPO	20 university students	Fully immersive VR	Log data	Internal	During training	Level 2
(Nykänen *et al.*, 2020)	Construction-related	Both RA and MPO	119 workers	Fully immersive VR	Questionnaire	External	1 month after training	Level 3
					Knowledge Test		Before and after training	Level 2
					Interview + Questionnaire		Before and after training	Level 1
(Shi *et al.*, 2020)	Construction-related	MPO	120 university students	Fully immersive VR and Non-immersive VR	Log data	Internal	During training	Level 2
					Questionnaire	External	After training	Level 1
(Wang *et al.*, 2020)	Construction-related	MPO	32 university students	Fully immersive VR	Log data	Internal	During training	Level 2
					Questionnaire	External	After training	Level 1
(Serpa *et al.*, 2020)	Manufacturing and assembly related	MPO	13 operators	Non-immersive VR	Log data	Internal	During training	Level 2
					Questionnaire	External	After training	Level 1
(Numfu *et al.*, 2020)	Manufacturing and assembly related	MPO	27 university students	Fully immersive VR	Questionnaire	External	After training	Level 1
(Vaquero-álvarez *et al.*, 2020)	Clinical or dental-related	Both RA and MPO	80 students	Non-immersive VR	Knowledge test	External	After training	Level 2
					Questionnaire	External	After training	Level 1
(Mondragón -Bernal, 2020)	Manufacturing and assembly related	Both RA and MPO	28 participants	Semi-immersive VR	Questionnaire	External	After training	Level 1
(Sakowitz *et al.*, 2020)	Clinical or dental-related	MPO	28 participants	Fully immersive VR	Knowledge test	External	Before and After training	Level 2
(Jain *et al.*, 2020)	Clinical or dental-related	MPO	4 participants	Fully immersive VR	Questionnaire	External	After training	Level 1
(Diego-Mas *et al.*, 2020)	Manufacturing and assembly related	RA	70 workers	Fully immersive VR	Questionnaire	External	3 month after training	Level 3
					Knowledge Test		After training and after 3 months	Level 2
					Interview + Questionnaire		After training	Level 1
(Ahn *et al.*, 2020)	Construction-related	RA	189 workers	Non-immersive VR	Knowledge Test	External	Before and after training	Level 2
			54 managers		Questionnaire		After training	Level 1
(Pedro *et al.*, 2020)	Construction-related	RA	45 participants (educators, safety experts, and university students)	Non-immersive VR	Interview + Questionnaire	External	After training	Level 1
(Osti *et al.*, 2021)	Construction-related	MPO	20 university students	Fully immersive VR	Log data	Internal	During training	Level 2
					Knowledge Test	External	Before and after training	
					Questionnaire		After training	Level 1
(Adami *et al.*, 2021)	Construction-related	Both RA and MPO	50 construction workers	Fully immersive VR	Knowledge test	External	Before and after training	Level 2
(Hernández-Chávez *et al.*, 2021)	Manufacturing and assembly related	MPO	20 university students	Fully immersive VR and Non-immersive VR	Questionnaire	External	After training	Level 1
(Kazar & Comu, 2021)	Construction-related	RA	42 university students	Non-immersive VR	Multiple Choice	External	Before, after, and 2 weeks after training	Level 2
(Dhalmahapatra *et al.*, 2021)	Construction-related	Both RA and MPO	19 operators	Fully immersive VR and Non-immersive VR	Log data	Internal	During training	Level 2
					Questionnaire	External	After training	Level 1
(Stransky *et al.*, 2021)	Chemical Process / Laboratory related	RA	290 university students	Non-immersive VR	Knowledge test	External	Before and after training	Level 2
(Bernard *et al.*, 2021)	Aviation-related	RA	112 students	Non-immersive VR and Semi-immersive VR	Questionnaire	External	After training	Level 1
(S. Li *et al.*, 2021)	Clinical or dental-related	MPO	36 participants	Fully immersive VR and Non-immersive VR	Questionnaire + interview	External	After training	Level 1
(Rahouti *et al.*, 2021)	Clinical or dental-related	RA	78 participants staff from hospital	Non-immersive VR	Knowledge test	External	Before and After training	Level 2
					Questionnaire	External	Before and After training	Level 1
(Joshi *et al.*, 2021)	Construction-related	RA	32 university students	Fully immersive VR	Knowledge test	External	Before and after training	Level 2
					Questionnaire	External	After training	Level 1

(Song *et al.*, 2021)	Construction-related	MPO	108 high school students	Fully immersive VR	Questionnaire	External	Before and after training	Level 1
(Y. Han *et al.*, 2021)	Construction-related	RA	40 participants	Fully immersive VR	Log data	Internal	During training	Level 2
					Questionnaire	External	After training	Level 1
(Poyade *et al.*, 2021)	Chemical Process / Laboratory related	Both RA and MPO	28 participants	Fully immersive VR	Knowledge test	External	After training	Level 2
					Questionnaire	External	After training	Level 1
(Xu & Zheng, 2021)	Construction-related	RA	10 participants	Fully immersive VR	Log data	Internal	During training	Level 2
					Questionnaire	External	After training	Level 1
(Kwegyir-Afful *et al.*, 2021)	Chemical Process / Laboratory related	RA	54 university students	Fully immersive VR	Questionnaire	External	After training	Level 1
(Grandi *et al.*, 2021)	Manufacturing and assembly related	MPO	5 participants	Fully immersive VR	Log data	Internal	During training	Level 2
					Questionnaire	External	After training	Level 1
(Beh *et al.*, 2021)	Construction-related	MPO	15 participants	Fully immersive VR	Log data	Internal	During training	Level 2
					Questionnaire	External	After training	Level 1

2.4. Findings and discussion

2.4.1. What topics have researchers investigated for VR-based health and safety training in various high-risk industries?

Of the 59 papers analysed, VR-based technologies have been used for H&S training in the following industrial sectors comprising high-risk activities: construction (n=25), clinical or dental (n=9), manufacturing and assembly (n=9), chemical process/laboratory (n=7), aviation (n=3), mining (n=2), electric power and electronics (n=2), maritime (n=1), and agricultural (n=1). This range of industries is due to the potential of VR technologies to create digital analogues for real-life scenarios that can be used for training, including both normal and abnormal operating conditions, in which stress drivers can still be incorporated while ensuring a safe training setting (Bissonnette *et al.*, 2019; Dholakiya *et al.*, 2019). Specifically, the use of VR-based technologies in the field of construction was evident when compared to other industries (Figure 2.4). One of the possible reasons why majority of the researchers used VR-based technologies in construction industries is its high accident and fatality rates (Pedro *et al.*, 2020). Thus, there is a need of a more engaging and interesting training environment which effectively trains them to exercise better safety practices. Another possible reason may be the ease of simulation and development of VR in construction related activities compared to other high-risk industries.

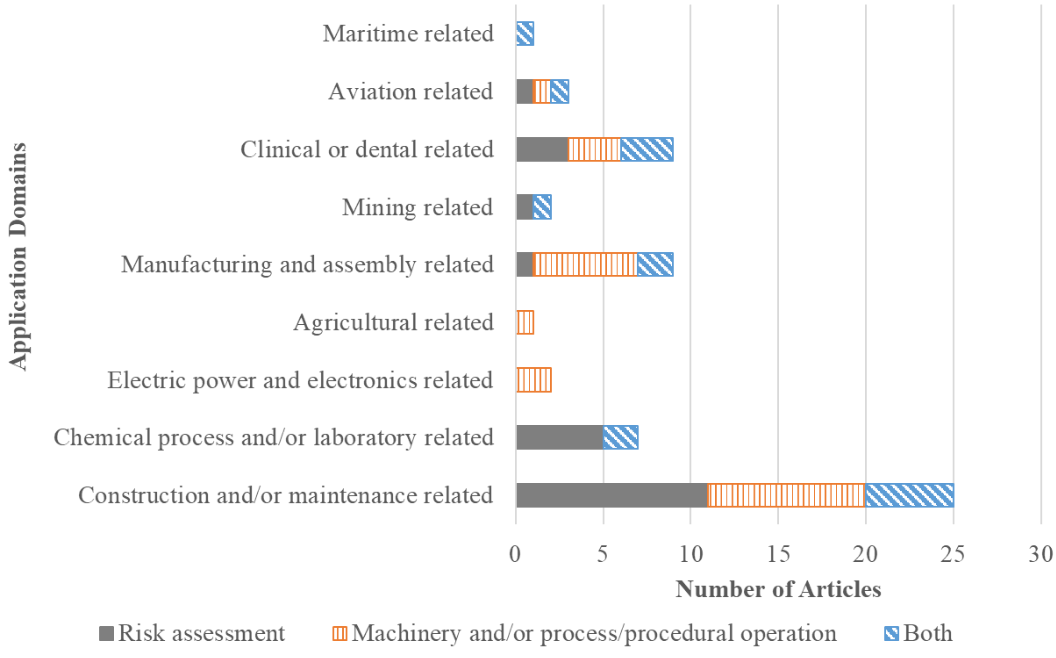


Figure 2.4. Distribution of the reviewed articles based on application domains.

In terms of the specific H&S topics taught in VR-based training, 22 out of 59 articles (Figure 2.4) reported in this literature review used VR technologies to upskill trainees on how to assess risk(s) in various high-risk industries. Risk assessment is the process of assessing the nature and likelihood of undesirable effects that may occur following exposure to hazards (*e.g.* biological, chemical, or physical) in a systematic way (Brecher, 1997). Most manuals and instructions on different safety training (*e.g.* construction, chemical, mining) list hazard identification, risk analysis and evaluation, risk control, and risk assessment documentation and review as the important steps (Health and Safety Executive, 2014). The first step is the identification of hazard(s) which requires learners to investigate and determine how and when a hazardous situation can lead to a certain accident(s). Risk analysis and evaluation is the second step which requires learners to understand the nature of the identified hazard(s) and determine the impact of corresponding risk(s). The third step is the risk control which requires learners to implement appropriate action(s) to identified risk(s) and the last step is the risk assessment documentation and review which requires learners to keep a formal record of the risk assessment.

Despite the consistent adaption of the abovementioned steps, the safety performance of the trainee group will remain low if the training programme is unengaging and passive. As the literature indicates, numerous hazards remain unrecognised and poorly managed in several high-risk industries such as construction-related workplaces, due to adoption of sub-standard practices and delivery methods in training programmes (Jeelani *et al.*, 2020).

To bridge these gaps, several researchers adapted and used VR-based training methods for risk assessment training such as the abovementioned list of risk assessment steps in various high-risk industries such as in construction, chemical process/laboratory, clinical or dental, aviation, manufacturing and assembly (Diego-Mas *et al.*, 2020), and mining (Isleyen & Duzgun, 2019), as these technologies can recreate a realistic but safe 3D environment of some hazardous workplace scenarios where the trainees can improve their risk assessment skills through the learning-by-doing approach (Table 2.4. and Figure 2.5). For instance, the study of Han *et al.* (2021) used a VR wearable device (HTC Vive) to locate, analyse, and mitigate hazards such as structural collapse, injuries by heavy equipment and injuries by manual handling or lifting at construction sites. Moreover, Kwegyir-Afful and his colleagues (2021) also used VR wearable device (HTC Vive) to recognise, evaluate, and control the fire hazard at a gas power plant.

Table 2.4. Distribution list of the authors based on the health and safety topic (risk assessment steps).

Application Domains	Authors
Construction	Ahn *et al.*, 2020; Albert *et al.*, 2014; Y. Han *et al.*, 2021; Joshi *et al.*, 2021; Kazar & Comu, 2021; Lin *et al.*, 2011, 2018; Pedro *et al.*, 2020; Perlman *et al.*, 2014; Pham *et al.*, 2019; Xu & Zheng, 2021
Chemical process/laboratory	Kwegyir-Afful *et al.*, 2021; Nazir *et al.*, 2015; Nicoletti & Padovano, 2019; Stransky *et al.*, 2021; Tawadrous *et al.*, 2017
Clinical or dental	Dorozhkin *et al.*, 2017; Polivka *et al.*, 2019; Rahouti *et al.*, 2021
Aviation	Bernard *et al.*, 2021
Manufacturing and assembly	Diego-Mas *et al.*, 2020
Mining	Isleyen & Duzgun, 2019

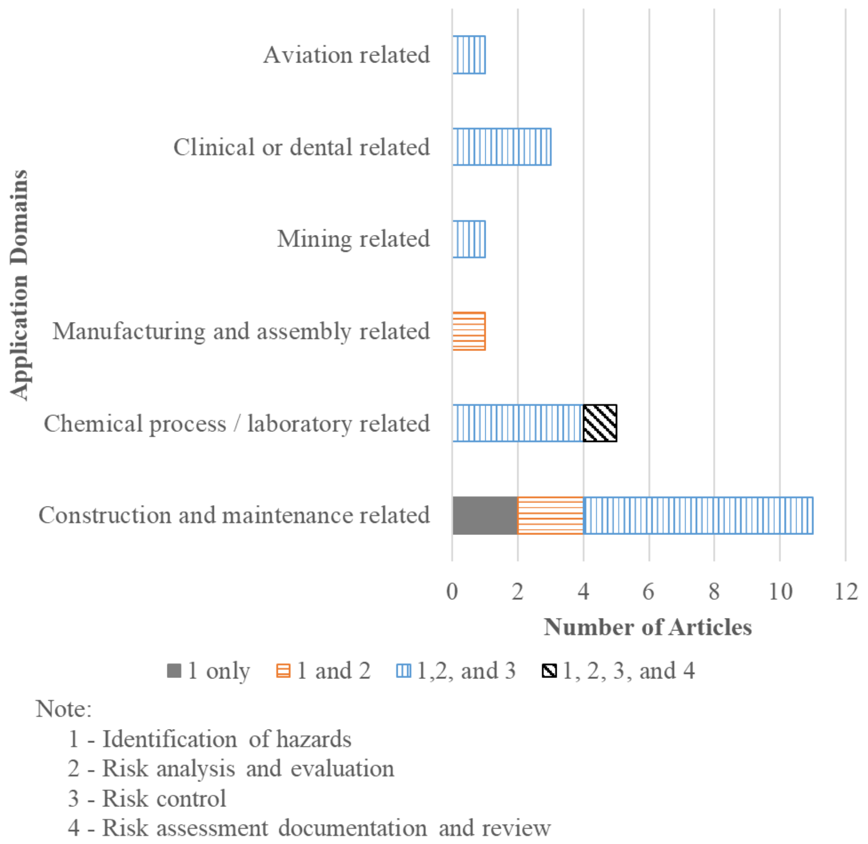


Figure 2.5. Distribution of the reviewed articles based on the health and safety topic (risk assessment steps).

Another 22 out of 59 papers (Figure 2.4) reported in this literature review used the VR-based technologies to train learners on how to control and manipulate machinery and/or equipment process/procedural operation. It is important to apply appropriate processes in different equipment and machinery in an inherently dangerous environment. The development of this practical skill requires an on-site experiences in a sustained period of time. Thus, it is important for the new employees to have on-site hands-on practice with the equipment and/or machines involved so they can fully appreciate and lessen the corresponding risks involved (Serpa *et al.*, 2020). However, it is often impractical to carry out training on actual machinery and/or equipment process /procedural operation safety training on the workplace when this interrupts on-site operations. As an alternative, employees are usually provided with a set of guidelines in the form of two-dimensional (2D) pictures and text which covers topics from terminology up to the operation and maintenance of equipment and/or machines.

Unfortunately, safety training delivering information through the abovementioned procedure usually offers a low level of engagement, presence, as well as realism since it is difficult for the new employees to fully visualize and understand the information provided from 2D pictures and text (Numfu *et al.*, 2020). To bridge these gaps, several researchers developed

VR-based machinery and/or equipment process/procedural operation safety training in construction, manufacturing and assembly, electric power and electronics, aviation, and agriculture as shown in Table 2.5 and Figure 2.6. For instance, Dado *et al.* (2018), used VR wearable device (HTC Vive) to allow trainees to use and become familiar with operation of an industrial lathe. Moreover, Song *et al.* (2021), used HTC Vive to train users in the operation of different cranes (*e.g.* overhead crane, tower crane, and container crane) by providing a virtual experience on how to operate these different type of cranes. As confirmed by the studies of the abovementioned researchers, adapting the VR-based technologies allows trainees to safely study and practice the operating procedures of the given machine/equipment that closely resemble the real environment they will encounter on-site.

Table 2.5. Distribution list of the authors based on the health and safety topic (machinery and/or equipment process/procedural operation).

Application Domains	Authors
Construction	Beh *et al.*, 2021; Choi *et al.*, 2020; Guo *et al.*, 2012; H. Li *et al.*, 2012; Osti *et al.*, 2021; Shi *et al.*, 2020; Song *et al.*, 2021; Vahdatikhaki *et al.*, 2019; Wang *et al.*, 2020
Manufacturing and assembly	Dado *et al.*, 2018; Gallegos-Nieto *et al.*, 2017; Grandi *et al.*, 2021; Hernández-Chávez *et al.*, 2021; Numfu *et al.*, 2020; Serpa *et al.*, 2020
Electric power and electronics	Ayala García *et al.*, 2016; Ogbuanya & Onele, 2018
Aviation	Tian *et al.*, 2015
Agriculture	Ojados Gonzalez *et al.*, 2017

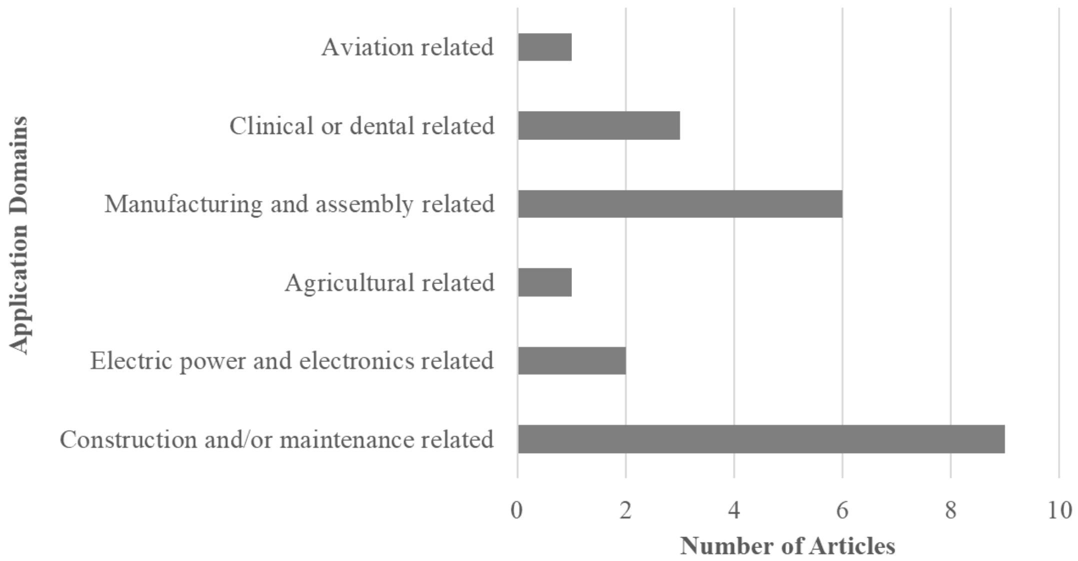


Figure 2.6. Distribution of the reviewed articles based on the health and safety topic (machinery and/or equipment process/procedural operation).

Lastly, some authors (Figure 2.4) used VR technologies to train users on both risk assessment and machinery and/or equipment process/procedural operation in construction, clinical or dental, chemical process/laboratory, manufacturing and assembly, aviation, maritime, and mining as shown in Table 2.6 and Figure 2.7. For instance, Dhalmahapatra *et al.* (2021) used Oculus Rift to help the learners to grasp the sequence of overhead crane operations as well as the process of managing the possible hazards while working.

Table 2.6. Distribution list of the authors based on the health and safety topic (both topics).

Application Domains	Authors
Construction	Adami *et al.*, 2021; Dhalmahapatra *et al.*, 2021; Le *et al.*, 2015; Nykänen *et al.*, 2020; Sacks *et al.*, 2013
Clinical or dental	Beyer-Berjot *et al.*, 2014; A. M. D. Jensen *et al.*, 2012; Vaquero-álvarez *et al.*, 2020
Chemical process/laboratory	Makransky *et al.*, 2019; Poyade *et al.*, 2021
Manufacturing and assembly	Leder *et al.*, 2019; Mondragón-Bernal, 2020
Aviation	Blickensderfer *et al.*, 2012
Maritime	Hjelmervik *et al.*, 2018
Mining	Liang *et al.*, 2019

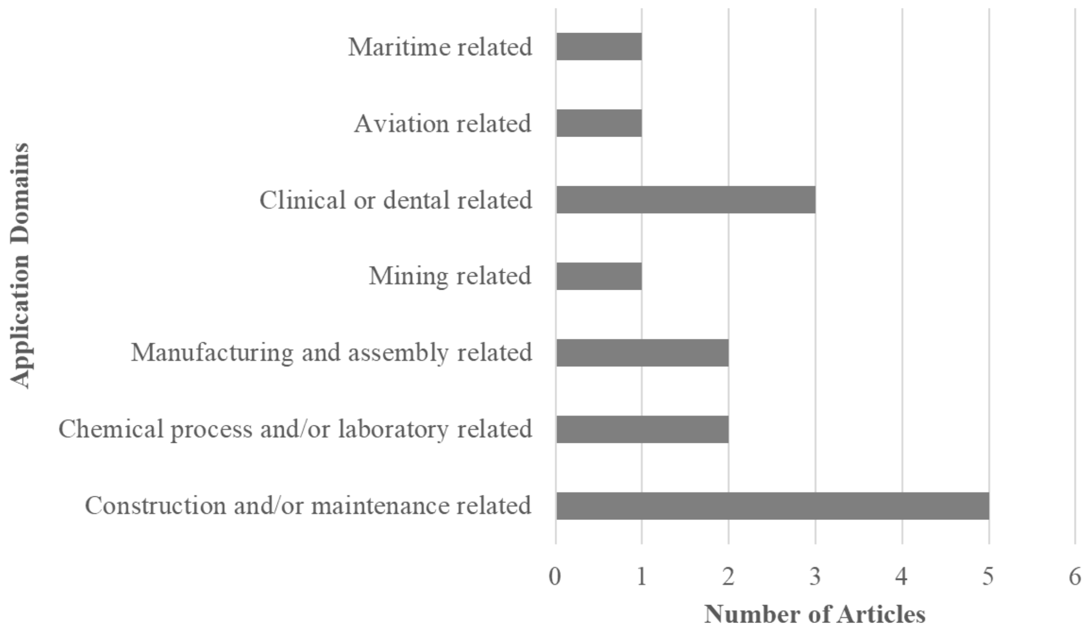


Figure 2.7. Distribution of the reviewed articles based on the health and safety topic (both topics).

2.4.2. What types of VR were used to deliver health and safety training in various high-risk industries?

Depending on the level of immersion, type of interactive and display device used, VR can be classified as either non-immersive (*i.e.* desktop), semi-immersive, or fully immersive VR (van Wyk & de Villiers, 2019). Non-immersive or desktop VR uses a conventional PC monitor, speakers and mouse to display virtual reality environment (VRE), sound, and interaction, respectively (van Wyk & de Villiers, 2019). On the other hand, semi-immersive or projected VR uses a system consisting of multiple projectors and projection screens, speakers, and controllers to display VRE, sound, and interaction, respectively while a fully immersive VR uses a head-mounted display (HMD) with earphones and motion tracking device to display VRE, sound, and interaction, respectively (van Wyk & de Villiers, 2019). Examples of non-immersive VR, semi-immersive VR, and fully immersive VR are shown in Figure 2.8.

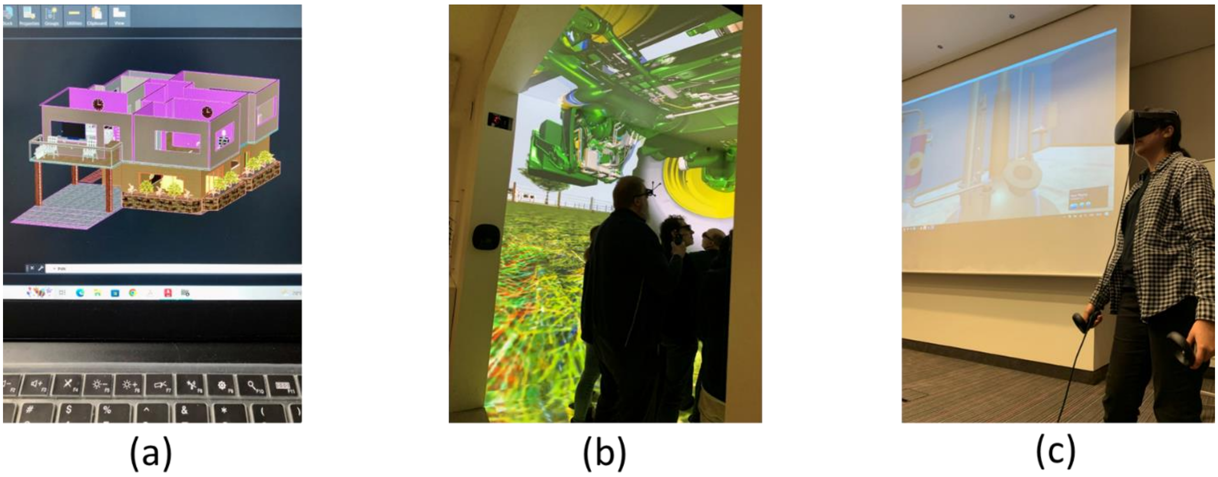


Figure 2.8. Example of (a) non-immersive VR, (b) semi-immersive VR, and (c) fully immersive VR.

In order to explore the use of different VR-based technologies used for H&S training in various high-risk industries, Figure 2.9 shows the distribution of VR technology types that were reported in the reviewed publications from 2011 to 2021.

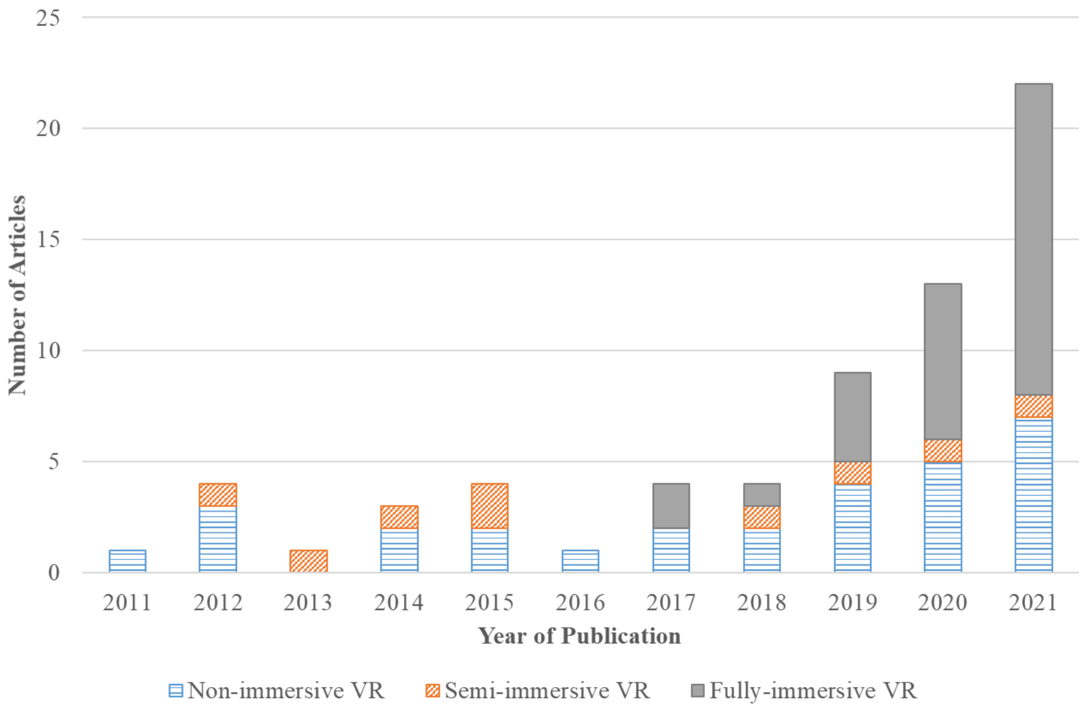


Figure 2.9. Distribution of VR technology types that were reported in the reviewed publications. (Note that a paper could potentially belong to multiple VR technology types.)

As shown in Figure 2.9, there are a few studies that used non-immersive VR technologies for H&S training in high-risk industries over the past 11 years. For instance, Lin *et al.* (2011), Ayala García *et al.* (2016), Nicoletti & Padovano, (2019), and Serpa *et al.* (2020) used non-immersive VR technologies, such as desktop computers, as a tool for H&S training in construction, electric power, chemical process, and manufacturing-related industries, respectively. Although Freina & Canessa (2015) noted, the non-immersive VR technologies lacks the feeling of presence (*i.e.* the subjective feeling of "being" in the task environment) compared to immersive VR and this leads to lower engagement and transfer of learning, their lower cost and a limited accessibility of the immersive type VR-based technologies means they are still used to some extent.

Compared to non-immersive VR, semi-immersive VR gives a greater sense of presence (An & Park, 2018). However, only nine authors (Blickensderfer *et al.* (2012), Sacks *et al.* (2013), Perlman *et al.* (2014), Nazir *et al.* (2015), Tian *et al.* (2015), Hjelmervik *et al.* (2018), Leder *et al.* (2019), Mondragón-Bernal (2020), and Bernard *et al.* (2021)) used semi-immersive type of VR technology such as the cave automatic virtual environment (CAVE) for H&S training in high-risk industries over the past 11 years (Figure 2.9). The low preference for semi-immersive VR compared to the other two types is due to the financial and management considerations. For instance, the construction and the installation of a new CAVE facility, consisting of a multiple high-resolution projectors and projection screens, is frequently complex, costly, and laborious in maintenance work if done by a single entity. (Havig *et al.*, 2011). Even if various entities collaborated in order to share the use of its facilities, only a limited number of institutions would still have access to the same. However, it should be pointed out that some researchers still prefer to use CAVE for H&S training in various high-risk industries as this technology has the ability to allow multiple participants to interact and share ideas/experiences with each other at the same time (Muhanna, 2015).

On the other hand, the number of publications on the use of fully immersive VR for health and safety training in various high-risk industries increased significantly from 2019 (Figure 2.9), and nowadays comprises the vast majority of the studies published. One of the reasons for this paradigm shift is the continuous improvement of these fully immersive VR over time. From the release of the first commercial VR head-mounted display (Oculus Rift) in 2013, the hardware and in display resolution have improved significantly over the last few years (L. Jensen & Konradsen, 2018). For instance, the typical field of view (FOV) of older HMDs was between 25 to 60 degrees, but new types of HMDs have FOVs above 100 degrees (Riva *et al.*, 2016). Another reason is the potential of the fully immersive technologies to provide a high degree of presence and immersion. Fully immersive VR allows users to be completely isolated from the real world, thus letting the user focus entirely on the VR environment to spend more time on the learning tasks, and gain better skills (L. Jensen & Konradsen, 2018).

The content of the training itself is highly important in evaluating its overall quality. However, a meritorious content is not sufficient for the effectiveness or training. Thus, in addition to the content, the quality of the training may also be significantly affected by the degree of user immersion. A high level of user immersion ensures better retention through simulation of real-life scenario (Ragan *et al.*, 2010). This is especially relevant in chemical plants wherein some types of technical know-how are better taught through practical application (Patle *et al.*, 2019).

Aside from the ability of these technologies to offer a better user experience, the significant reduction in the cost of the new generation of HMDs made these the best choice for several companies as well as research institutions. Given these benefits, it is expected that there will be a progressive increase in the number of publications on this type of VR for H&S training in various high-risk industries in the next few years.

2.4.3. What are the outcome(s) measured for establishing the effectiveness of the VR-based health and safety training in various high-risk industries?

The adoption of VR-based technologies for H&S training in high-risk industries has increased in the past 11 years. However, since the success of a given training method depends on the degree to which the training prepares trainees for real-world situations (*a.k.a.* transfer of training), it is important to understand how to analyse and evaluate the outcome(s) of a VR-based health and safety training in high-risk industries. Prior to objectively investigating the impact of the training, it is vital to categorise the outcome(s) of the training programme.

Most of the existing training evaluation models such as the Hamblin's model (1974), the organisational elements model (1995), the Indiana University model (1996), among others, follow the Kirkpatrick four-step model which was first developed in 1959 and was progressively updated over the last 40 years (Alsalamah & Callinan, 2021; Hirsh & Carter, 2002; Rafiq, 2015). According to Bates (2004), the popularity of Kirkpatrick model may be explained by several factors. Primarily, the model established a systematic way to understand training evaluation by providing terms with clear definitions to which training outcomes and other information may be identified. More importantly, Kirkpatrick's model developed a four-level categorization of training outcomes that conveys highly descriptive, yet straightforward information about each respective level. The appeal of Kirkpatrick's model also lies in its simplification of the complex process of training evaluation. In effect, the model streamlines an otherwise complex network of variables in the training process to come up with the training outcome evaluation (Bates, 2004).

The prevailing criticism on Kirkpatrick's model is its tendency to largely focus on the outcomes of the training, thus, neglecting other aspects of the learning procedure (Bushnell, 1990). However, Giarangco *et al.* (2010) contends that the abovementioned statement is inaccurate, as variables used in Kirkpatrick's model rely on specific and predefined learning objectives and training design. Indeed, contrary to the criticism of the model, application of Kirkpatrick model not only focuses on the overall training outcome, but also considers the effectiveness of training design, whether the objectives of each phase of the training have been achieved and the observable changes in perspective of the trainees, among others.

According to Kirkpatrick (2006), evaluating the outcome(s) of training can be classified into four levels (Figure 2.10). The first level is the reaction level where reactions of the trainees (*e.g.* trainees thoughts) are identified and the satisfaction of the trainees is measured. The second level, which can be described as the measurement of the increase in knowledge or intellectual capability as a consequence of the training is known as the learning level. The third level is the behaviour level. This requires measuring the change behaviour that transfers to actual performance in the job as a result of the training. The final level involves the assessment of the impact of training in terms of organizational outcomes (Kirkpatrick, 2006). Although the framework is usually applied in a step-by-step manner to map the process of evaluating the success of a given training, some training does not require the implementation of this step-by-step process. For instance, training such as information security training requires the trainees not only retain the information but also apply this information at work. Thus, an assessment designer should focus on levels 2 and 3. However, if the institution developed a new information security training method (*e.g.* VR-based training), then the assessment designer should apply the Kirkpatrick evaluation model in a step-by-step manner as they need to assess the overall impact of the newly developed training method and make a practical judgement whether to adopt or replace the existing training method.

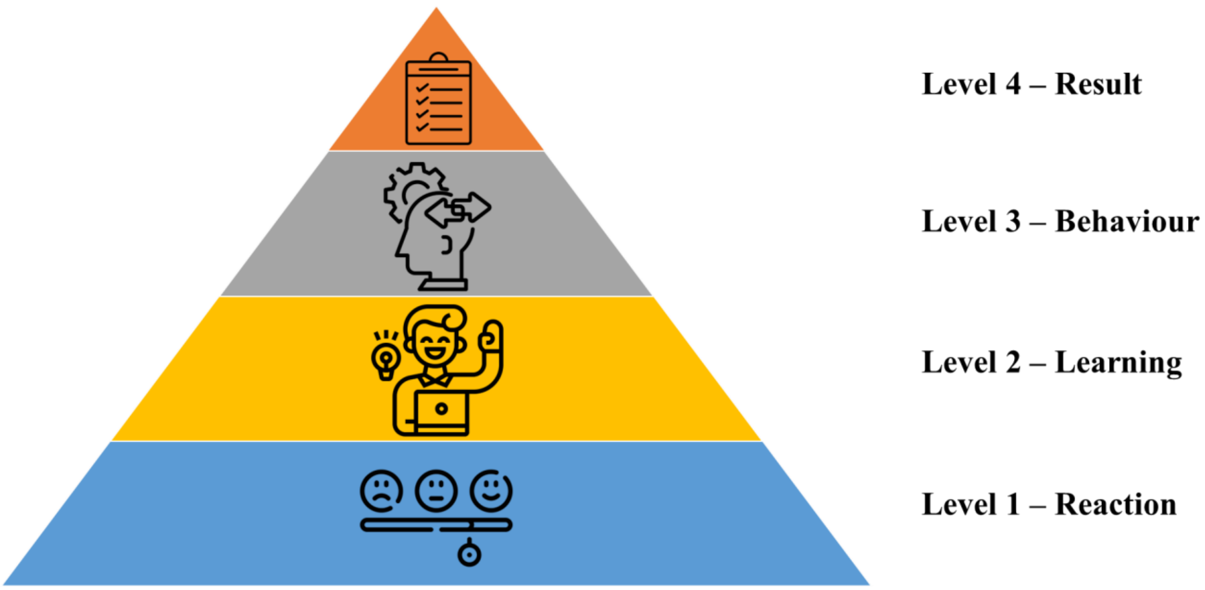


Figure 2.10. The Kirkpatrick's training evaluation model.

With regards to the implementation of VR-based technologies, it is important for the stakeholders need to consider the evidence from Kirkpatrick four level of training evaluation to decide whether or not to invest in VR-based training. Figure 2.11 shows the distribution of the outcome(s) measured for VR-based health and safety training based on Kirkpatrick's training evaluation model over a span of 2011 to 2021.

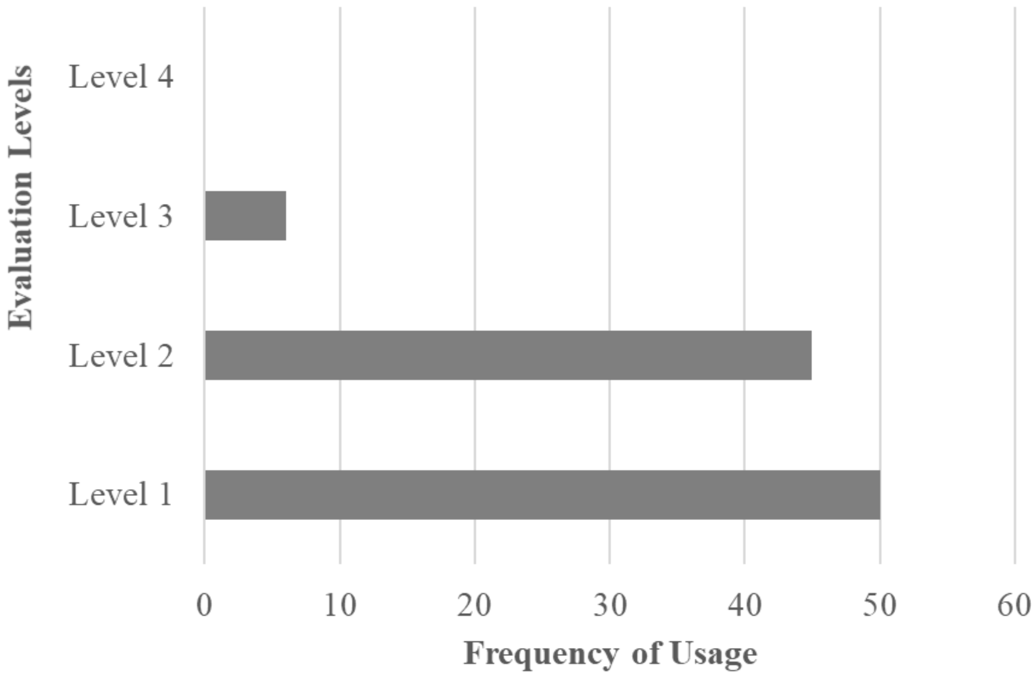


Figure 2.11. Distribution of the outcome(s) measured based on Kirkpatrick's training evaluation model.
(Note that a paper could potentially belong to multiple Kirkpatrick's evaluation level.)

As shown in Figure 2.11, a number of authors such as Lin *et al.* (2011), Isleyen and Duzgun (2019), and Numfu *et al.* (2020), evaluated the reaction level (Level 1) outcome for different type of VR-based H&S training. For instance, the *t*-test results of the study with 32 participants conducted by Joshi *et al.* (2021), suggest that the motivation levels achieved in fully immersive VR for precast/pre-stressed concrete industry safety training was higher than the group which was only provided by an instructional video as the former is more effective in making students learn from their mistakes in the VR environment. Although VR-based training has gained a significant level of attention in several high-risk industries such as in medical, in aviation, and even in engineering, most of the authors still conducted the reaction level training evaluation, either stand-alone or in conjunction with other training evaluation methods, for the past 11 years. The main reason why researchers still conduct evaluation on reaction level outcomes is because they want to verify the potential of these new technologies (regardless of the type) for a specific topic of H&S training in high-risk industries.

However, having a positive outcome at the reaction level does not guarantee that there is a knowledge/skills acquisition when VR-based technologies were adopted. This is because the data gathered from reaction level (level 1) only reflects the overall reaction/experience (*e.g.* satisfaction, enjoyment, *etc.*) of the given training. As a result, authors such as Nazir *et al.*

(2015), Tawadrous *et al.* (2017), and Choi *et al.* (2020) evaluated the learning level (Level 2) outcome for different type of VR-based H&S training (Figure 2.11). For instance, the t-test results of the study with 189 workers conducted by Ahn *et al.* (2020), suggest that regardless of the length of work experience, site workers trained through non-immersive VR construction safety training showed higher level of understanding than the group of workers who were training through traditional lecture setting ($t = 2.848$, $p < 0.05$ for more than nine years of working experience, $t = 2.237$, $p < 0.05$ for within three to eight years of working experience, and $t = 2.090$, $p < 0.05$ for less than 2 years of working experience).

Fewer authors used the third level of Kirkpatrick's model to evaluate the long-term effect of the given VR-based training on the behaviour of the trainees (Figure 2.11). For instance, the results of the generalised linear mixed modelling method performed by Nykänen *et al.* (2020) confirmed that there was a greater increase in the self-reported safety performance of the 119 participants (*e.g.* identifying factors affecting safety) one month after their VR-based construction safety training compared to lecture-based safety training (confidence interval = 95%, estimate = 0.46, $p < 0.05$). Moreover, there was no reported study measuring the fourth level of Kirkpatrick's model which evaluates the organisational results and the cost and return on investment of the training for the past 11 years. The low number of articles on the third and fourth level of Kirkpatrick's model is due to the fact that the process of measuring the extent of learning transferred to job behaviour (Level 3) or the overall success of the training (Level 4) requires researchers to perform longitudinal studies (*i.e.* conducting and reviewing pre-defined performance metrics in a pre-set time interval through observation). As most of the projects have limited funding duration, it is difficult to obtain additional funding if there is some delay due to unforeseen circumstances and this leads to fewer studies dealing with the behaviour and results level outcomes (Caruana *et al.*, 2015).

2.4.4. What assessment techniques were used to evaluate the outcome(s) of VR-based health and safety training in various high-risk industries?

Since it is important for every training to measure specific performance criteria that are essential to the development of skilled personnel, understanding appropriate assessment methods plays a vital role. Broadly speaking, assessment is a systematic process of recording and presenting information (*e.g.* knowledge, skills, *etc.*) about learner accomplishment and instructional processes (Brookhart, 1999). Figure 2.12 illustrates the assessment process.

Figure 2.12. Assessment process.

Generally, assessment can be categorised into two groups: summative (assessment of learning) and formative (assessment for learning) (Loh, 2012). Summative assessment analyses the understanding and mastery of the topic after an activity is completed while formative assessment makes use of regular interactive measurements that identify points of improvement for better learning outcomes (Loh, 2013; Sadler, 1989). The information for both summative and formative assessment can be collected through several ways such as conducting paper-and-pencil tests (*e.g.* multiple choice, matching, *etc.*), self/peer/supervisor-report (*e.g.* feedback, observation, *etc.*).

According to Loh (2011), many researchers claimed that formative assessment can have a positive effect on the learning processes as it provides continuous and timely assessment information which can point out, shape, and improve the specific area of difficulties trainees are having (Loh, 2011). Unfortunately, the implementation of formative assessment is not an easy task for most of the trainers and lecturers, especially teaching in large lecture classes, as they cannot afford extra time and effort to provide valuable feedback to address the gap between their present and their projected performance (Bennett, 2011). However, as the use of an online learning has evolved considerably for the past few years, the concept of a digital-based simulation or a game assessment (*i.e.* process of automated collection, organisation, documentation, and presentation of scores and the corresponding feedback on individual learner performance managed through the medium of digital devices such as computers, VR-HMDs, and *etc.*) becomes broadly recognised as a solution to the abovementioned implementation problem (Bulut *et al.*, 2019).

According to Eseryel *et al.* (2011), a digital-based simulation or a game-based assessment may be categorised as either external or internal assessment. Some examples of external assessment include interviews, multiple-choice questionnaires (MCQ), knowledge tests, and practical tests which are similar to the traditional assessment methods. On the other hand, data provided through the simulation and game files that log the actions of the player and game variables are examples of internal assessment. Both assessment methods can be used as summative or formative depending on the timing of implementation (*e.g.* before, during, and/or after playing the game/simulation). The main difference between these two types of assessment is that the external assessment is not normally part of the game/simulation course, and it will interrupt the game/simulation. On the other hand, the internal assessment is typically used in the game/simulation course without unnessesary interference with the game/simulation itself (Eseryel *et al.*, 2011). Figure 2.13 shows the distribution of the assessment methods used to measure the different outcomes of training evaluation based on Kirkpatrick's model.

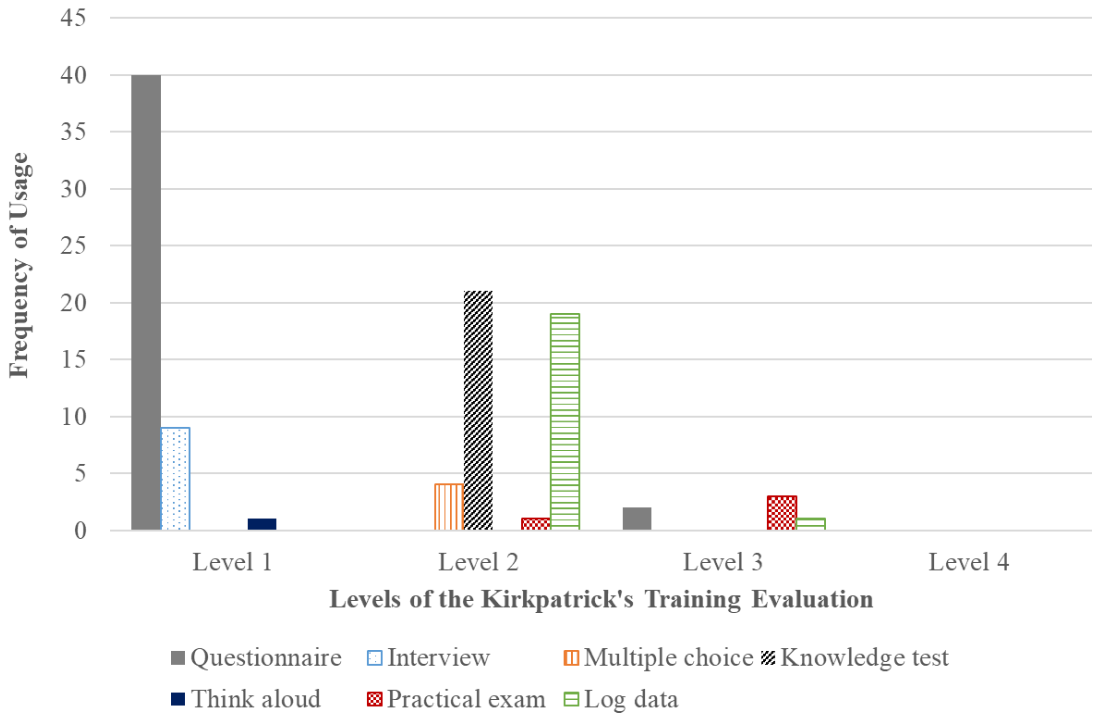


Figure 2.13. Distribution of the assessment methods used to measure the outcomes of training evaluation.
(Note that a paper could potentially belong to multiple Kirkpatrick's evaluation levels)

Most of the articles considered in this review used external assessment, such as self-assessment questionnaires (*e.g.* intrinsic motivation, perceived enjoyment, presence, self-efficacy, effectiveness and satisfaction questionnaires), or interviews, to evaluate the first level of Kirkpatrick's model (reaction criteria) of the trainees which measures their satisfaction with the VR-based H&S training (Figure 2.13). For instance, Wang *et al.* (2020) used a Likert scale questionnaire to determine the change in terms of the satisfaction of the trainees in undertaking either the fully immersive VR or traditional lecture-based scaffolding erection operation training in construction industry. Their results from a paired-sample test with 32 university students confirmed that compared to lecture-based training, participants who used VR-based training showed a stronger impact on satisfaction (5.81 vs. 6.81). Moreover, the VR-based training approach (mean value = 6.56) was more helpful compared to the lecture-based training (mean value = 5.75). One of the reasons for the frequent usage of external assessment over internal assessment for evaluating the first level of Kirkpatrick‘s model is because questionnaires are relatively easy to administer and implement. Another reason is because the implementation of internal assessment for measuring satisfaction/usability of VR-based training requires an additional work from the integration of the appropriate tools (*e.g.* emotion sensors) to the analysis of the desired variable(s) from the large amount of data corresponding to certain emotions (*i.e.* positive, negative, neutral) (Dzedzickis *et al.*, 2020).

As shown in Figure 2.13, both external and internal assessment methods can be used to evaluate how much knowledge/skills trainees gained in the VR-based training programme (Level 2 of Kirkpatrick's model). For instance, Ogbuanya and Onele (2018) used a knowledge test to assess the fundamental knowledge of electrical/electronic technology operation gained in non-immersive VR-based training compared to conventional classroom training. Using analysis of covariance (ANCOVA), their results indicate that virtual reality positively affected the academic performance of 142 learners as there was a significant differences in the knowledge test scores of the participants who used the non-immersive VR-based training (mean = 71.7) and the traditional method (mean = 60.1) (Ogbuanya & Onele, 2018). Although it is easier to evaluate the effectiveness of the VR training by checking the learning of the players using conventional assessment methods, assessment of trainees through external assessments, such as the paper-and-pencil format, is only efficient for measuring some simple outcomes, such as declarative knowledge, rather than the development of complex skills (*e.g.* problem-solving, teamwork and collaboration, *etc.*) (Garcia Fracaro *et al.*, 2021).

To maximize the potential of VR-based technologies, several authors used internal assessments such as log data to trace and capture learner-generated data (*e.g.* correct actions, tasks completed). For instance, Nazir *et al.* (2015) used log data to capture the actual performance of 24 participants on locating correct valves, opening or closing a valve, and/or identify leakages for safety training in a chemical plant. Their results showed that participants trained in a VR environment were able to identify more leakages (67%) and manually operated valves (83%) compared to conventional methods such as power point presentation (42% and 50%, respectively) (Nazir *et al.*, 2015). Through integrating internal assessment, it is possible to create an automated assessment capable of measuring complex skills, which translate into better performance in the real world (Shute & Wang, 2016). However, Loh (2012) argued that creating game/simulation-based analytics requires a lot of work from discovering useful metrics for measuring human performance, to verifying the corresponding equivalence of digital and actual actions, and to identifying strong predictors from thousands of information points available in each data set. Nevertheless, it is expected that there will be a progressive increase in the usage of internal assessment for H&S training in various high-risk industries in the next few years given the rapid advancement in the fields of data mining and machine learning, which will facilitate the development of the log data analysis.

Figure 2.13 further shows even fewer authors evaluating the third level of Kirkpatrick's model compared to the previous two levels as the former requires significant amount of time and money. Moreover, it was evident from the figure that most of the authors prefer using external assessment methods, such as questionnaires and practical exams, to internal methods (*e.g.* log data) for evaluating the amount of learning transferred to job behaviour. For instance, Makaransky *et al.* (2019) used practical exams, a situational judgement scenario, to assess 105 university students on the amount of learning transferred to job behaviour in a chemical laboratory setting after training using a VR-based platform. The results from the Dunnett's test showed that the students in the fully immersive VR-based safety training showed better performance to demonstrate appropriate laboratory skills and behaviour in the practical tests compared to conventional text-based safety manual training (p = 0.031, d = 0.58) (Makransky *et al.*, 2019). On the other hand, Albert *et al.*, (2014) used longitudinal collection and analysis of log data to measure the behavioural criteria of around 15-36 trainees in VR-based construction safety. Their study confirmed that participants were able to increase their hazard recognition skills from 46% to 77% in the post-intervention phase and maintained this score until the end of the 16th working period. They also stated that it was important to have a support from funding agencies and partnership with a wide range of industry professionals with varied skills and experience in order to capture the needed variables from the log data and accurately measure patterns of change that can be used for the determination of behavioural criteria (Albert *et al.*, 2014).

2.4.5. Does VR-based health and safety training in various high-risk industries have the potential to improve the training evaluation outcome(s) compared to traditional and/or other VR-based training methods?

VR technology creates a representation of real-life scenario which allows trainees to be exposed and to be trained in dealing with hazardous situations within a safe 3D setting. In this context, several authors explore the potential impact of such VR-based H&S training to improve the outcome(s), measured based on Kirkpatrick's training evaluation model, compared to traditional (*e.g.* lecture, PowerPoint presentation, audio-visual presentation, *etc.*) and/or other

VR-based training methods (Dhalmahapatra *et al.*, 2021; Makransky *et al.*, 2019; Osti *et al.*, 2021). At present, there is a lack of research on whether VR as a mode of active learning is more effective than other modes of active learning (*e.g.* on-the-job training) when it comes to H&S training. This is probably due to the difficulty of implementing other modes of active learning in scenarios involving H&S, such as incidents in the chemical plant involving explosions, corrosions, emission of toxic substances, and the like. Thus, while VR-based training generally provides a safer environment, its effectiveness *vis-à-vis* other modes of active learning has not been sufficiently tested. One study suggests that training in a virtual reality environment produces similar results as compared to a real-life environment such as pilot plant (Garcia Fracaro *et al.*, 2020). However, it must be noted that this study only considers the procedural aspect of the chemical plant operations and not situations involving accidents or crises. This question may be researched further to provide insights whether active learning *per se* is actually the cause of the potential improvement of the training outcomes and not the VR application itself.

Figure 2.14 shows the distribution of the studies which compare the outcome(s), measured based on Kirkpatrick's training evaluation models, between different types of VR-based training or between VR-based training *vis-à-vis* traditional training. There are 59 articles considered in this study, 50 of which discuss reaction level (Level 1) while 45 and 6 of which discuss learning level (Level 2) and behaviour level (Level 3), respectively.

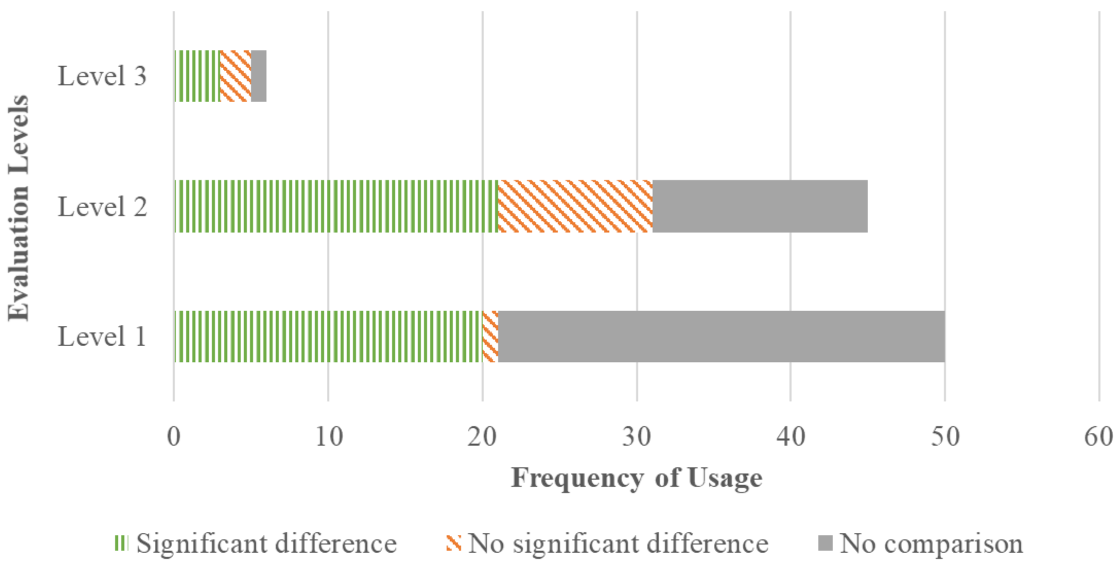


Figure 2.14. Distribution of the studies which compare the outcome(s) between different types of VR-based training or between VR-based training *vis-à-vis* traditional training.
(Note that a paper could potentially belong to multiple Kirkpatrick's evaluation levels.)

As shown in Figure 2.14, 21 out of 50 papers compared the reaction level between different types of VR-based training or between VR-based training *vis-à-vis* traditional training in various high-risk engineering industries. Among the 21 papers, the results of a statistical test (*e.g. t*-test, ANOVA, *etc.*) of the 17 studies showed that the VR-based training provides a greater degree of reaction level compared to the traditional setting (Ahn *et al.*, 2020; Beh *et al.*, 2021; Blickensderfer *et al.*, 2012; Diego-Mas *et al.*, 2020; Guo *et al.*, 2012; Joshi *et al.*, 2021; Leder *et al.*, 2019; H. Li *et al.*, 2012; Makransky *et al.*, 2019; Nykänen *et al.*, 2020; Pham *et al.*, 2019; Polivka *et al.*, 2019; Poyade *et al.*, 2021; Rahouti *et al.*, 2021; Sacks *et al.*, 2013; Tian *et al.*, 2015; Xu & Zheng, 2021). For instance, Leder *et al.* (2019), in a study with 68 apprentices, stated that compared to the traditional training method (PowerPoint presentation), there was a better degree of immersion and presence in the semi-immersive VR (CAVE) environment. Moreover, three of the papers compared the reaction level of different types of VR-based training system. For instance, Hernández-Chávez *et al.* (2021), Dhalmahapatra *et al.* (2021), and Li, S. *et al.* (2021) compared the reaction level between fully immersive VR and non-immersive VR. Their results showed that the fully immersive VR was better with respect to several reaction level criteria such as the ease of operation, the ease of learning, realism, immersion, presence, and/or graphics quality compared to desktop VR. On the other hand, Osti *et al.* (2021), in a research study with 20 participants, showed that there was no statistical difference in the usability scores between the fully immersive VR training and traditional video training. However, it is important to note that the usability score of fully immersive VR was higher than the score of traditional video training (Osti *et al.*, 2021). As most of the scores for the reaction level were better in the VR-based setting, this suggests that the use of VR-based training in various high-risk engineering industries may have a higher potential to provide enhanced degree of reaction level compared to the traditional setting.

31 out of 45 papers considered in this study compared the learning level between different types of VR-based training or between VR-based training *vis-à-vis* traditional training in various high-risk engineering industries (Figure 2.14). Out of the 45 papers, 20 papers showed that the VR-based training provides higher learning and/or performance scores with respect to several H&S topics, such as risk assessment and/or machinery and process operation compared to the traditional setting (Adami *et al.*, 2021; Ahn *et al.*, 2020; Ayala García *et al.*, 2016; Blickensderfer *et al.*, 2012; Dado *et al.*, 2018; Diego-Mas *et al.*, 2020; Gallegos-Nieto *et al.*, 2017; Kazar & Comu, 2021; H. Li *et al.*, 2012; Liang *et al.*, 2019; Nazir *et al.*, 2015; Nykänen *et al.*, 2020; Ogbuanya & Onele, 2018; Perlman *et al.*, 2014; Pham *et al.*, 2019; Rahouti *et al.*, 2021; Sacks *et al.*, 2013; Shi *et al.*, 2020; Stransky *et al.*, 2021; Tian *et al.*, 2015). For instance, Pham *et al.* (2019), in a study with 40 university students, stated that users who

used non-immersive VR obtained higher scores (mean = 80.1%) in hazard investigation compared to users who used the traditional lecture-based platform (mean = 76.3%). Moreover, Dhalmahapatra *et al.* (2021), in research with 19 operators, compared the degree of learning level between fully immersive VR and non-immersive VR and their *t*-test results showed that the safety performance of the users trained in the fully immersive VR was better than the performance of the users trained in the non-immersive VR. On the other hand, researchers such as Makransky *et al.* (2019), Leder *et al.* (2019), Polivka *et al.* (2019), Sakowitz *et al.* (2020), 2 case studies of Osti *et al.* (2021), Joshi *et al.* (2021), Poyade *et al.* (2021), and Beh *et al.* (2021), showed in their respective studies that there was no statistical difference in the performance scores between the VR-based training and traditional training methods. Although the results confirmed that there was no statistical difference, six studies stated that performance achieved through the VR-based training was higher than using the traditional training method. Majority of the studies imply favourable use of VR-based training as it can provide higher learning and performance scores compared to traditional training methods. This indicates that the use of VR-based training may have a higher potential to provide better learning and/or performance to the users.

In terms of the behaviour level, five out of six papers compare the level between VR-based training and traditional training in various high-risk engineering industries (Figure 2.14). Out of the five papers, researchers such as Ayala García *et al.* (2016), Makransky *et al.* (2019), and Nykänen *et al.* (2020), confirmed that there was a significant difference in terms of the ability of the users to demonstrate appropriate skills and behaviour in VR-based training compared to traditional training methods. On the other hand, authors such as Leder *et al.* (2019), and Diego-Mas *et al.* (2020), confirmed that there was no significant difference in the behaviour level in their respective studies. However, it is important to note that the behaviour performance scores of the VR-based training for both studies were higher than the traditional training. This suggests that the use of VR-based training in various high-risk engineering industries may have a higher potential to provide improved degree of behaviour level compared to the traditional setting.

2.5. Implications

Although this chapter focuses on the use of VR-based H&S training in various high-risk industries, researchers and stakeholders may consider the findings of this study as a basis for proposing a training design framework that may be adopted to align the VR-based training with desired training outcome and assessment method. Figure 2.15 shows the proposed training design framework.

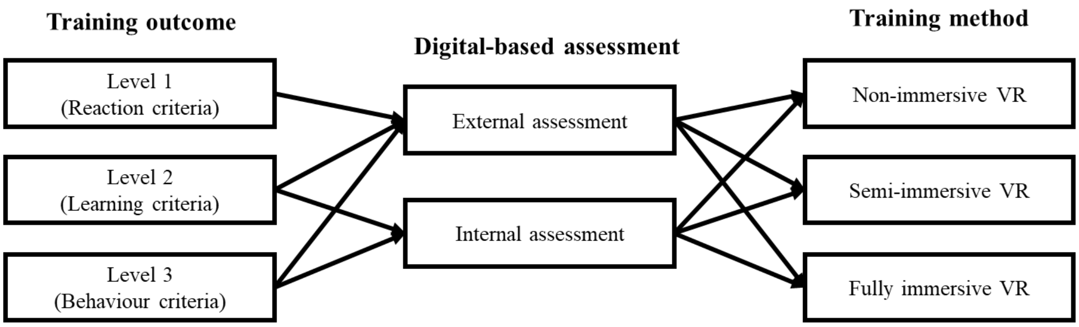


Figure 2.15. Proposed training design framework.

As training outcomes tend to focus on what trainees should achieve upon completion of a certain training programme, practitioners must clearly and accurately define these training outcome(s). For instance, if the training is newly developed, it might be beneficial for the stakeholders to initially assess the usability and satisfaction of the new training (Level 1) as well as the immediate knowledge/skills gain (Level 2). After analysing those outcomes, a decision can be made on whether to continue and whether to invest/develop the training programme through evaluating the behavioural change among trainees (Level 3).

After defining and determining the desired outcome(s), it is important to choose the appropriate digital-based assessment method(s) for evaluating the outcome(s). For instance, in evaluating the training satisfaction (Level 1), it is better to use external assessment methods such as questionnaires or interviews as these are proven to be easier to implement. On the other hand, both level 2 and level 3 outcomes can be evaluated by internal and external assessment methods. For instance, if the institution wants to create and develop an automated assessment, it might be beneficial to consider internal assessment methods rather than external assessment methods. However, practitioners must consider the advantages as well as the disadvantages of every assessment type as this will affect the required resources (*e.g.* human, financial, and time) needed by the institution/organisation.

Upon aligning the desired training outcome(s) and the assessment method(s), practitioners can then select the suitable training method that will boost the engagement of the trainees. For instance, if the institution/organisation is aiming to create an affordable, realistic, but safe replica of specific dangerous training activity, it might be appropriate to consider fully immersive VR as a training tool.

2.6. Conclusions

This study presents a review of the existing articles relating to the use of VR-based H&S training in various high-risk industries. It also provides some insights on the types of VR, topics of H&S training, and types of assessment techniques and training evaluation. In addition, this study explored the potential of VR-based H&S training to improve the training evaluation outcome(s) compared to traditional and/or other VR-based training methods. 59 articles reporting specific assessment techniques were considered and analysed. The results indicated that most of the industries used VR-based technologies to train users in risk assessment, machinery and/or equipment process/procedural operation, or both topics. Moreover, the use of fully immersive VR increased rapidly due to the recent improvements in hardware, display resolution, and price. In terms of the outcomes measured for establishing effectiveness of the VR-based H&S training, the interest of the trainers is focused on the measurement of the amount of change in the satisfaction and/or learning achievement of trainees within a short span of time. For instance, most of the researchers were using external assessments such as questionnaires, and interviews for training satisfaction studies as these are proven to be easier to implement. Moreover, external assessment such as knowledge tests and MCQs were also used to evaluate the amount of declarative knowledge gained by the trainees in the VR-based training. On the other hand, some researchers used internal assessment methods, such as logs of data, to create an automated assessment which is capable of measuring complex skills such as problem solving and teamwork. Lastly, the VR-based H&S training was also found to have the potential to improve the reaction level, learning level, and behaviour level compared to traditional training methods.

In conclusion, the findings from this study can contribute to and support the practitioners and safety managers in practice by providing a training design framework that may be adopted to align the VR-based training with desired training outcome and assessment method. As discussed (for examples, refer to Chapter 2.5), the training design framework should identify the appropriate digital-based assessment and its platform in consideration of the desired learning outcome, with a view of ensuring that the trainees acquire mastery of their functions in the shortest possible period. This study can also be used as a basis to suggest that researchers should consider conducting more research on the evaluation on the effectiveness of VR-based H&S training focusing on the third and fourth level of Kirkpatrick's model using internal assessment method such as log files.

2.7. Note

Part of this chapter has been published in the **International Journal of Educational Technology in Higher Education**: Toyoda, R., Russo Abegão, F., and Glassey, J. (2022). VR-based health and safety training in various high-risk engineering industries: a literature review. International Journal of Educational Technology in Higher Education, 19(1), 42. https://doi.org/10.1186/s41239-022-00349-3

Chapter 3. Drivers of immersive virtual reality adoption intention: a multi-group analysis in chemical industry settings

Chapter 2 highlighted the rapid advancement and growing market of immersive virtual reality (IVR) based health and safety (H&S) training in various high-risk industries such as aviation, construction, and chemical industries. Although research on the feasibility, reliability, and ease of use of IVR has received considerable attention, little is known about the specific factors contributing to the decision to adopt IVR in the H&S training setting. Since the success of implementing IVR in H&S training depends, amongst other factors, on the individuals willing to use it, this chapter uses the modified Unified Theory of Acceptance & Use of Technology 2 (UTAUT 2), which reviewed and integrated the best aspects of eight prominent models to develop its improved framework. This unified model is used to examine the effect of factors such as performance expectancy (PE), effort expectancy (EE), social influence (SI), and hedonic motivation (HM) that may motivate individuals to adopt IVR-based technology into their training. In addition, this study performs a multi-group analysis based on nationality, prior IVR experience, and/or length of work experience, to analyse the potential similarities and/or differences in perception and acceptance towards using IVR-based technology. The quantitative research data were gathered using an online questionnaire from 438 chemical operators and/or employees who either speak German, French, or English. Partial least squares structural equation modelling (PLS-SEM) and multi-group analysis (MGA) based on SmartPLS^{TM} version 3 was used to carry out the path and multi-group analyses. The results show that the behavioural intention (BI) towards the adoption of IVR was influenced by performance expectancy, effort expectancy, and hedonic motivation for all participating subpopulations. However, the relationship of social influence to behavioural intention was not supported for respondents with prior IVR experience and for respondents coming from western regions. Although Henseler's-based multi-group PLS analysis reveals that there was no significant difference between the group comparisons, it is still important to take into account these socio-demographic factors as there are definite group differences in terms of the ranking order of each construct for the IVR adoption intentions among each subpopulation. The implications are also discussed.

3.1. Introduction

Since several companies around the world are adapting and embracing the concept of industry 4.0, technologies such as virtual reality (VR) technology gained a significant level of attention and created a paradigm shift in several areas of training in the fields of chemical (Feise & Schaer, 2021), medical (Bissonnette *et al.*, 2019), and aviation industries (Clifford *et al.*, 2019). As pointed out by many researchers, training materials such as PowerPoint presentations or pre-recorded lectures only provide and explain instructions and rules without realistic feeling for the given scenarios (Arkorful & Abaidoo, 2015). Such approaches are not particularly effective, especially in the abovementioned fields (Dholakiya *et al.*, 2019). For instance, scenarios involving risky situations at work are difficult to simulate realistically due to cost, safety, and environmental implications (Manca *et al.*, 2013). As VR technology, particularly immersive virtual technology (IVR) such as the IVR head-mounted displays and the cave automatic virtual environment (CAVE), can provide users with a safe 3D training environment space, promoting knowledge acquisition through active involvement, it has become possible to create a representation of real-life scenario for training under normal or abnormal situations within a safe setting while retaining stress drivers (Bissonnette *et al.*, 2019; Dholakiya *et al.*, 2019).

The use of IVR technology in chemical industry setting can improve higher-order thinking competencies that are important for scenario-based training, such as problem-solving and communication skills. For instance, Colombo *et al.* (2014) used an IVR-based training scenario of responding to pressurised liquid butane (C4) leakage due to inadvertent excavator operation. Their study showed that participants trained in an IVR environment performed 50% better in fault diagnosis, which develops problem-solving skills through deeper situational awareness, than those trained with conventional slide-supported presentations (Colombo *et al.*, 2014). Nevertheless, it is still necessary to investigate the perceptions and acceptance of users towards the application of IVR technologies in chemical industries. This is due to the fact that this specific area scarcely uses the applied theories on technology acceptance. The results of this investigation will provide answers to the question whether it is, at present, appropriate to use IVR or not, and to further develop the IVR technology that would best fit the intended uses. Since the success rate of employing IVR technology is dependent on the number of people who are eager to try and use this technology, it is vital to identify the factors affecting IVR adoption intentions (Van Slyke *et al.*, 2007). For this purpose, the Unified Theory of Acceptance and Use of Technology 2 (UTAUT 2) model, an extension of the UTAUT model, will be used in this study. Both the UTAUT and UTAUT 2 models proved to be more comprehensive and give higher explanatory power than other models, as validated by Venkatesh *et al.*, in 2003 and 2012, respectively. Prominent models for evaluating technology acceptance include Technology Acceptance Model (TAM), Theory of Planned Behaviour (TPB), Model of PC Utilisation

(MPCU). Nonetheless, the UTAUT and UTAUT 2 model already reviewed, compared, improved, and consolidated eight popular models including the abovementioned ones. More importantly, the main reason for choosing the UTAUT 2 model is that it facilitates the understanding of the adoption and use intention of the consumers (*i.e.* employees and operators) towards IVR technology in a range of industries, including chemical processes. The UTAUT 2 model considers seven key factors: performance expectancy (PE), effort expectancy (EE), social influence (SI), facilitating conditions (FC), hedonic motivation (HM), price value (PV), and habit (H). The model also includes three moderators: age, gender, and experience. Both the key factors and the moderators are considered to affect the behavioural intention (BI) and/or use behaviour (USE). In this model, all the seven key factors affect the behavioural intention, while the key factors FC, H, and BI influence the use of behaviour.

Previous studies investigated the acceptance of IVR in various fields. For instance, Hartl and Berger (2017) explored the consumer acceptance of VR glasses in entertainment content (*e.g.* watching 360° documentary video, and playing 3-minutes game) using the extended UTAUT 2 model. In their study, researchers recruited a total of 155 participants (53% male, 75% without prior experience of IVR, and average age of 24 years) from a public university in Germany. They found that only 3 out of 6 factors, PE, SI, and H, showed significant effects on behavioural intention to adopt IVR system (Hartl & Berger, 2017). On the other hand, Kunz and Santomier (2019) used the extended UTAUT 2 model to investigate the acceptance of VR technology in a sports context. A total of 570 participants were recruited from a university in southern Germany (67% male, 61% without prior experience of IVR, and average age of 23 years). They learned that only 3 out of 7 factors, PE, SI, and HM, showed significant effects on behavioural intention to adopt IVR system (Kunz & Santomier, 2019). As observed, the IVR adoption intention results from both studies are somewhat different, although they both identified PE and SI as significant factors. The results obtained from these IVR adoption studies cannot be used to generalise to other groups as it may cause misinterpretation due to the differences in terms of sample demographics and fields of application of the technology. Hence, this study employs a modified UTAUT 2 model to investigate the factors for the IVR adoption from the perspective of operators and employees in the chemical industry.

Sarstedt *et al.* (2011) proposed that data analysis in technology acceptance research should not be limited to a single population (*i.e.* homogeneous representation of all observations) as interpreting the results may result in misleading conclusions. Thus, they proposed to apply a multi-group analysis (MGA) into partial least square (PLS) path model to ascertain whether there are relevant similarities and/or differences across groups. As such, this study uses partial least square multi-group analysis (PLS-MGA) to test the similarities and differences between the different sub-categories of chemical operators and employees in terms of their intentions to adopt IVR. The findings from this model will prove useful for understanding the acceptance of IVR among operators and employees after considering their background. The findings will enable chemical industries to implement more effective training programmes in futures, and to more judiciously consider investments in IVR technology for health and safety training.

3.2. Theoretical underpinning and hypothesis development

Although the awareness and the popularity of using new technologies like e-learning and VR have increased especially during the COVID-19 period, some people are still hesitant to adopt these new technologies. In order for the developers to understand and address the hesitance of people towards the adoption of new technology, it is important to investigate the interrelationship between influential factors and the behavioural intention. Without understanding the gap between what people claim through their attitudes and involvement and how they behave, the rate of success in implementing IVR training will be low. Given the relevance and the growing importance of IVR adoption intention in chemical industry, it is timely to examine the perceptions of operators and employees in terms of their intentions of IVR adoption for H&S training.

Researchers usually evaluates the differences in attitudes across various groups to determine the acceptability of a new technology in an organisation (Marques *et al.*, 2011). Acceptability may be ascertained by calculating the differences among attitude scores, but there should be a validation whether the criteria used in measuring attitudes are observed in the same way by the respondent groups (Cronbach, 1992; Cronbach & Furby, 1970).

Moreover, other underlying factors that affects the intention of the individual to accept and use a technology must be considered (Marques *et al.*, 2011; Sharma & Kumar, 2012). These factors may include beliefs, characteristics, and other external factors. However, it is often difficult to particularly measures these factors resulting to the researchers having to decide the most appropriate model in accordance with the specific circumstances (Venkatesh *et al.*, 2003). Thus, it is important to understand the different models for technology acceptance.

The first model, originated from social psychology, used to study acceptance of technology is the Theory of Reasoned Action (TRA) (Rondan-Cataluña *et al.*, 2015). According to this model, subjective norm, perceived behavioural control, and attitude toward the behaviour determine behavioural intention (*i.e.* the intent to carry out certain behaviour) (Fishbein &

Ajzen, 1975). Since TRA is not designed for assessing specific behaviour or technology, it can be applied to different fields (Rondan-Cataluña *et al.*, 2015).

To understand user acceptance in the field of information technology (IT), Davis adopted TRA and developed a Technology Acceptance Model (TAM) (Davis, 1989). This model suggests that perceived usefulness (PU) and perceived ease of use (PEOU) are the two main factors that affect the acceptance of a technology in the workplace (Rondan-Cataluña *et al.*, 2015). Over time, many tests verified the strong determinant effect of PU and PEOU on BI, so the TAM model was extended to TAM2 and TAM 3. The former model is based on the expansion of the antecedents of PU whilst TAM3 is based on the expansion of the antecedents of PEOU (Venkatesh & Bala, 2008; Venkatesh & Davis, 2000). TAM models have been widely used in a variety of technologies especially to website applications (Rondan-Cataluña *et al.*, 2015).

Since understanding the explanation of the use and the acceptance of a new technology has become one of the leading aspects of research in the IT area, many additional models have been devised (Rondan-Cataluña *et al.*, 2015). Among these models, several studies use Unified Theory of Acceptance and Use of Technology (UTAUT), which was formulated to explain and predict user acceptance and/or use of new technology (Bracq *et al.*, 2019; Ogourtsova *et al.*, 2019). Since there is a need to have an collective view of user acceptance to avoid mixing concepts of various theories, UTAUT model is synthesised from the measurable/verifiable comparison of eight models: the Theory of Reasoned Action (TRA), the Technology Acceptance Model (TAM), the Motivational Model (MM), the Theory of Planned Behaviour (TPB), the Combined TAM and TPB (C-TAM-TPB), the Model of PC Utilization (MPCU), the Innovation Diffusion Theory (IDT), and the Social Cognitive Theory (SCT) (Venkatesh *et al.*, 2003). The UTAUT model was validated using within-subject longitudinal data from different organizations (J. P. Li & Kishore, 2006; Venkatesh *et al.*, 2003). The model explained 77% and 55% of variance in behavioural intention to use a new technology and in technology use, respectively, as indicated in the empirical studies of acceptance of new software application for employees working in financial or customer services (Venkatesh *et al.*, 2003, 2016). Given that the abovementioned R^2 values were greater than the 40% and 30% recommended values for behavioural intention and use behaviour respectively (Zhou *et al.*, 2021), the UTAUT model demonstrates good explanatory power in the context of employees working in financial or customer services adopting new software application.

According to Venkatesh *et al.* (2003), the UTAUT model considers four key factors: performance expectancy (PE), effort expectancy (EE), social influence (SI), and facilitating conditions (FC) for explaining and predicting user acceptance and/or use of new technology.

The model also includes four moderators: age, gender, voluntariness, and experience. Both the key factors and the moderators are considered to affect the behavioural intention (BI) and/or use behaviour (USE). In this model, the key factors PE, EE, and SI affect the behavioural intention, while the key factors FC and BI influence the use behaviour (Venkatesh *et al.,* 2003). The relationships mentioned above is shown in Figure 3.1.

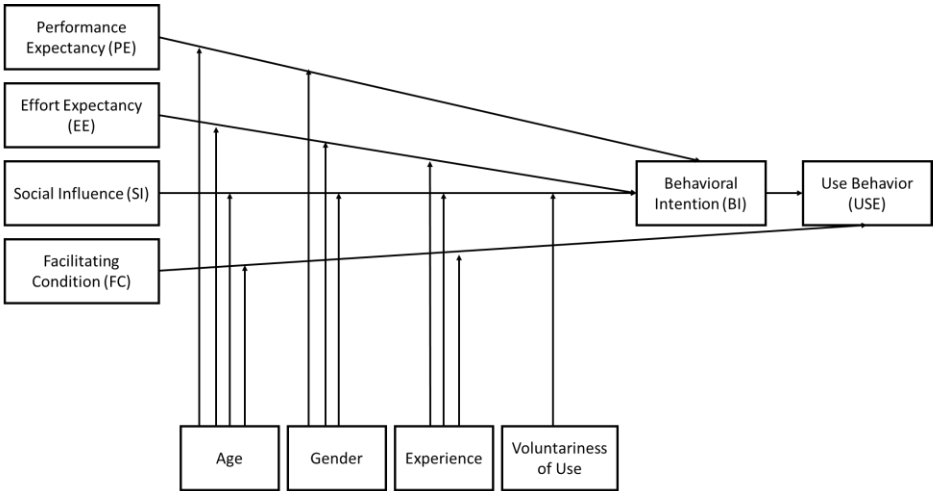
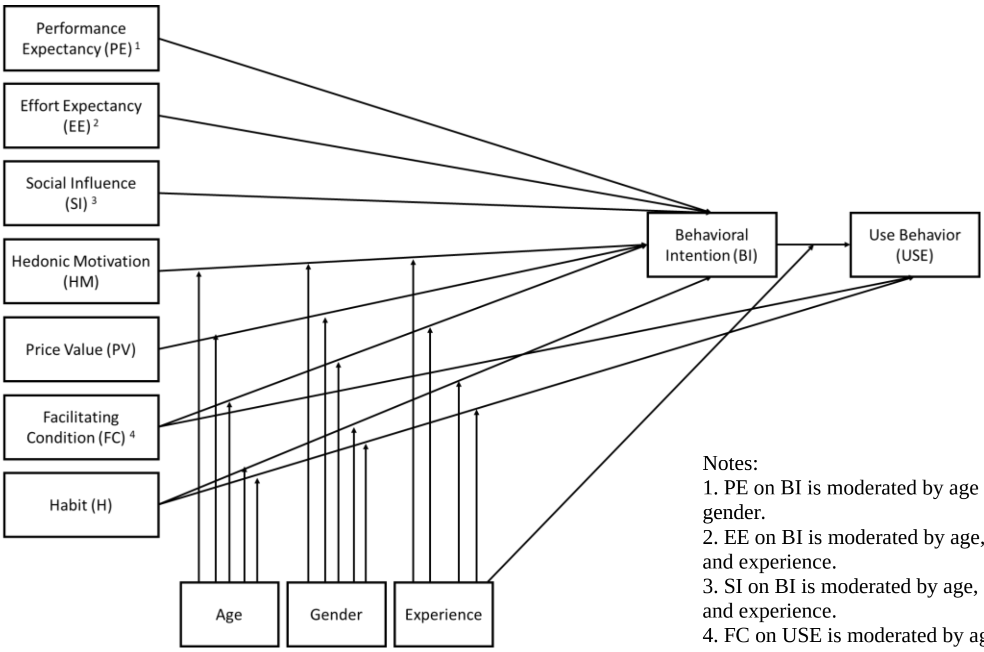


Figure 3.1. UTAUT Model (Venkatesh *et al.*, 2003).

Moreover, Venkatesh *et al.* (2012) established UTAUT2 model to address a new context of consumers. This model employed hedonic motivation (HM), price value (PV), habit (H), and facilitating condition (FC) as additional key factors that influence BI compared to the original UTAUT model. Additionally, habit also influences USE (Venkatesh *et al.*, 2012). UTAUT2 has been reported to be able to explain about 74% and 52% of variance in behavioural intention of use of new technology and in technology use for consumers, respectively (Venkatesh *et al.*, 2016). Given that the abovementioned R^2 values were greater than the 40% and 30% recommended values for behavioural intention and use behaviour respectively (Zhou *et al.*, 2021), the UTAUT2 model demonstrates good explanatory power in the context of consumers adopting new technology. These relationships are shown in Figure 3.2.



Notes:
1. PE on BI is moderated by age and gender.
2. EE on BI is moderated by age, gender, and experience.
3. SI on BI is moderated by age, gender, and experience.
4. FC on USE is moderated by age and experience.

Figure 3.2. UTAUT2 Model (Venkatesh *et al.*, 2012).

Since the adoption intention in an organizational context and the continuous usage intention in consumer context are explained in UTAUT and UTAUT2 research models, respectively, these baseline models are usually used to analyse the factors influencing individual intentions at different stages of technology adoption and use (Kupfer *et al.*, 2016; Rondan-Cataluña *et al.*, 2015). For instance, several studies confirmed that the abovementioned key factors have a significant influence on BI to adopt new technology, such as an e-scooter VR service (Huang, 2020), and a VR-based surgical simulator for scrub nurses (Bracq *et al.*, 2019). Thus, this study has employed the UTAUT 2, an adaptation of UTAUT to a consumer context, to explain and predict behavioural intention of users towards IVR adoption.

As this study focused purely on the perception of IVR adoption from the perspective of operators and employees in the chemical industry, the construct of PE has been conceptualised as the extent to which chemical operators and employees perceive IVR as a tool that would lead to additional improvement in their job performance. The construct of EE has been conceptualised as the extent to which chemical operators and employees perceive IVR to be simple to operate and easy to use. The construct of SI has been conceptualised as the extent to which chemical operators and employees perceive the expectations of their peers on their use of IVR. Finally, the construct of HM has been conceptualised as the extent to which chemical operators and employees perceive IVR as a tool to bring additional joy and enjoyment. From these, the following hypotheses, adapted and modified from Venkatesh *et al.* (2003, 2012), are proposed:

H1: *PE has a significant influence on BI to adopt IVR for H&S learning.*

H2: *EE has a significant influence on BI to adopt IVR for H&S learning.*

H3: *SI has a significant influence on BI to adopt IVR for H&S learning.*

H4: *HM has a significant influence on BI to adopt IVR for H&S learning.*

Several authors used the UTAUT model to identify the effect of different constructs towards IVR adoption intention using a single population sample. For instance, Shen *et al.* (2019) investigated the direct indicators affecting the intention of 376 university students (62.2% female, 54.8% undergraduate students, and average age of 22 years) in Taiwan to use VR-based head-mounted displays (HMDs) in learning through UTAUT model. They found that all four constructs (PE, EE, SI, and FC) of the UTAUT showed a positive and significant effect on the behavioural intention of the students to use HMDs in their learning activities (Shen *et al.*, 2019). However, according to Cheah *et al.* (2020), interpretation of results from a homogenous population could be misleading as all individuals have their own perceptions and evaluations of outcomes. Thus, they proposed to assess data by adding more subgroups of data into the model to minimize misinterpretation of results (Cheah *et al.*, 2020).

In relation to this proposal, a number of authors considered socio-demographic variables such as age, gender, nationality, and experience in an analysis of the behaviour of technology users. For instance, Venkatesh and his co-workers clustered participants based on their age, gender, and experience and analysed the willingness of users to accept and use new technology in the workplace, and the willingness of consumers to accept and use mobile internet technology using UTAUT and UTAUT 2 model, respectively. Both of these studies confirmed that socio-demographic variables were key factors in the BI to adopt and/or use new technology. On the other hand, Palau-Saumell *et al.* (2019) employed the extended UTAUT 2 model to compare the usage intention with mobile application for restaurant searches and/or reservations, and Ramirez-Correa *et al.* (2015), used the Technology Acceptance Model (TAM) to compare e-learning intentions. These authors confirmed that some of the socio-demographic variables (*i.e.* gender and age) were not key factors in the BI of users to adopt and/or use these technologies. Since there is some discrepancy among previous studies on the effect of socio-demographic variables based on the technology used, it is important to establish the effect of these socio-demographic variables on the behaviour of users toward IVR adoption. By doing so for different groups, the influence of the PE, EE, SI, and HM in terms of IVR adoption intention can be compared. Thus, the following hypotheses are proposed:

H5: *The influence of adoption in IVR is different between groups based on the nationality of the employees.*

H6: *The influence of adoption in IVR is different between groups based on the prior IVR experience of the employees.*

H7: *The influence of adoption in IVR is different between groups based on the length of work experience of the employees.*

Finally, conceptual model of IVR adoption intention modified from UTAUT 2 model is postulated as shown in Figure 3.3.

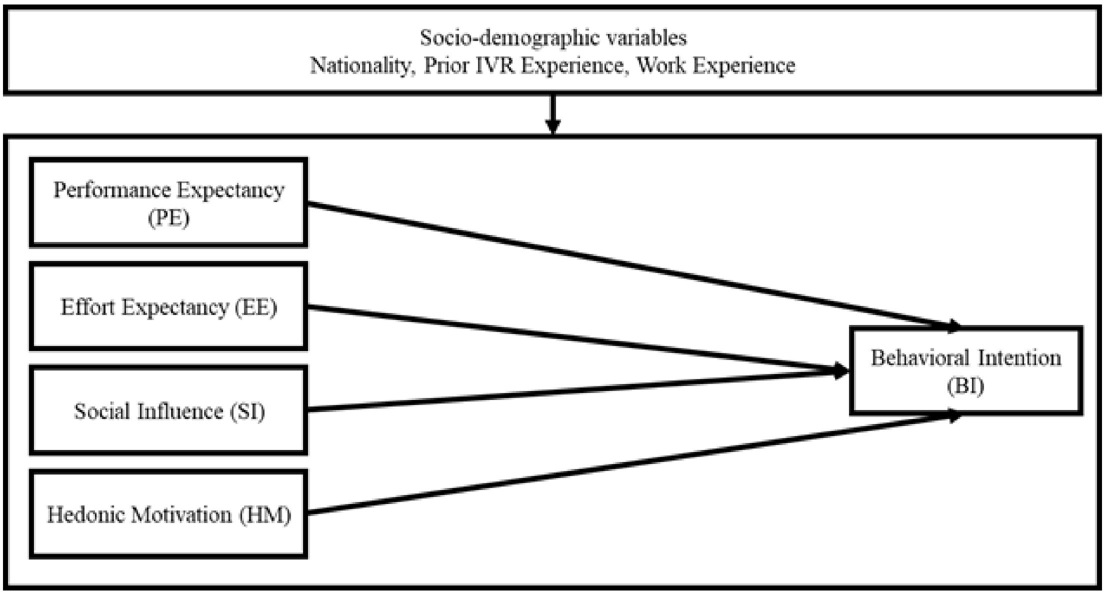


Figure 3.3. The modified UTAUT 2 model

3.2. Research Methodology

3.2.1. Questionnaire Design

To measure the perceptions of chemical industry employees toward H&S training using IVR technology, this study employed an online questionnaire survey. The final online questionnaire comprised of two sections. The first section covered the socio-demographic background of the respondents, including their gender, age, nationality, prior IVR experience, and the length of employment. The second section contained questions about PE, EE, SI, HM, and BI that were adopted from previously reported research using the UTAUT 2 model and that were verified as valid and reliable (Venkatesh *et al.*, 2012). For each item, some of the words were modified to better fit the scope of IVR games in training. The respondents indicated their agreement with each item on a six-point Likert scale, ranging from 1 for strongly disagree to 6 for strongly agree. The reason for choosing the 6-point Likert scale is that it gives a higher trend of discrimination and reliability as compared to a 5-point Likert scale (Chomeya, 2010).

Since this study also aimed to examine the difference in IVR adoption between eastern and western countries, questionnaires in English, French, and German were prepared. The questionnaire was originally created in English and was subsequently translated into French and German by a native speaker. A separate native speaker then performed a blind back-translation of the questionnaire into English, which was compared with the initial English version to ensure the uniformity and validity of the translation (Dorer, 2012). The English list of items used in the study with its corresponding constructs is shown in Table 3.1.

Table 3.1. Lists of measurement items used in the study.

Latent Variable	Item	Explanation
Performance Expectancy (PE)	PE_1	I think that using the VR environment will be useful for practicing H&S procedures.
	PE_2	Using VR environment will probably enable me to learn the H&S procedures more quickly.
	PE_3	If I use this VR environment, I will improve my performance on H&S procedures.
Effort Expectancy (EE)	EE_1	I think using the VR environment will be clear and understandable.
	EE_2	I think that it will be easy for me to operate the platform in which the VR environment is running.
	EE_3	I think that it will take too long to learn how to use the VR environment to make it worth the effort.*
Social Influence (SI)	SI_1	I think that the organization will support me in learning how to use the VR environment.
	SI_2	People who influence my behaviour at work think that I should use this VR environment.
	SI_3	I think my supervisor will be very supportive of the use of this VR environment for my job.
Hedonic Motivation (HM)	HM_1	I feel that it will be a bad idea to use the VR environment for H&S training.*
	HM_2	I think that the actual process of using the VR environment for H&S training is fun.
	HM_3	I think that using VR environment for H&S training will be very frustrating.*
Behavioural Intention (BI)	BI_1	If made available to me, I would recommend using the VR environment for learning to apply the H&S procedures to my colleagues.
	BI_2	If made available to me, I plan to continue to use VR environment for H&S training frequently.
	BI_3	I think that after using the VR for H&S training, I will be ready to use this learning environment for another training course.

Note: * - inverted item

3.2.2. Data Collection Process

Before the data collection, the questionnaire was tested by an academic expert in the field (Dr Bert Slof of Utrecht University) to ensure face validity as well as content validity. The questionnaire was revised according to the expert feedback and conducted a pilot readability test with postgraduate chemical engineering students and volunteer engineers in the chemical industry. Prior to the distribution of the questionnaire, an ethical authorisation was obtained from the Ethics Committee at the university.

The responses were collected from employees working in chemical industries situated in Europe and Asia. Since it was impossible to include all employees working in the country in the sample, convenience sampling, a non-probability sampling method was used. Four hundred and thirty-eight (438) completed questionnaires were collected and were subjected to data screening to eliminate invalid questionnaires. Since no incomplete or duplicated data were present, all responses were used for data analysis.

The power analysis using G*Power 3 analysis software was performed to ensure the sufficiency of the sample size (Faul *et al.*, 2007). From the calculation, the minimum required sample size for this study is 129. Thus, the sample of 438 collected data used in this study is sufficient.

3.2.3. Data Analysis

Since multivariate analysis involves the application of statistical methods that simultaneously analyse multiple variables, researchers need to select an appropriate method based on the underlying research question and the empirical data available (Hair *et al.*, 2017). Multivariate analysis can be categorized into two groups: first-generation (*e.g.* ANOVA, Multiple Regression, Factor analysis) and second-generation (*e.g.* structural equation modelling). Many researchers, especially in the field of social science, use the latter method since it enables simultaneous analysis of the hypothesized relationships in a given model, and also possible correlations between multiple dependent and independent variable (Hair *et al.*, 2017).

Given the abovementioned advantages, Structural Equation Modelling (SEM), a second-generation multivariate analysis was used in this study. There are two types of SEM: covariance-based SEM (CB-SEM) which is primarily used for theory testing and confirmation and partial least square SEM (PLS-SEM) which is often used for theory development from exploratory study or extension of an existing structural theory. Rather than confirming an established theory, this study intended to explore the modified version of the well-known technology acceptance theory (UTAUT2 model). Hence, it is more appropriate to use PLS-SEM. It requires a lower sample size and non-normally distributed data compared to covariance-based (CB) SEM (Hair *et al.*, 2017). Therefore, PLS-SEM using SmartPLS™ version 3 software was used to assess the measurement model and to test the path relationship between the constructs of the collected data (Ringle *et al.*, 2015).

Given that the aim of this study is to examine the differences in hypothetical relationships between groups, a multi-group analysis approach (MGA) in PLS-SEM was carried out. The first step is the creation of data groups based on the categorical variable of interest (Matthews, 2017). The next step involves the analysis of the measurement invariance (*i.e.* equivalence) of composite models (MICOM) across two or more groups following a three-step procedure: (1) configural invariance, (2) compositional invariance, and (3) the equality of composite mean values and variances (Henseler *et al.*, 2016). After establishing measurement invariance, a comparison of path coefficients among groups using Henseler PLS-MGA procedure was carried out to determine the significant differences between groups (Matthews, 2017).

3.3. Results

3.3.1. Participant Profile

The demographic information of the participants is summarised in Table 3.2. Out of the 438 participants, those coming from Asia account for 32.9% of the group compared to the participants coming from Europe representing 67.1% of the group. Males account for 66.4% of the participants, while females account for 33.6%. The majority of the participants are between the ages of 20-39 (57.3%) and have more than five (5) years of working experience (62.8%). Finally, more than 70% of the participants had prior experience in playing video games, but only 35.4% of them have tried head-mounted display VR.

Table 3.2. Demographic information of participants (n = 438).

Characteristics	Items	Frequency	Percentage
Nationality	Eastern countries	144	32.9
	Western countries	294	67.1
Gender	Male	291	66.4
	Female	147	33.6
Age	20-29	155	35.4
	30-39	96	21.9
	40-49	83	18.9

	50-59	88	20.1
	60 and above	16	3.7
Working Experience	Less than a year	37	8.4
	1-5 years	126	28.8
	6-20 years	153	34.9
	More than 20 years	122	27.9
Experience to VR	Yes	155	35.4
	No	283	64.6
Experience to Video Game	Yes	326	74.4
	No	112	25.6

3.3.2. Assessment of the measurement model

To use the scores obtained from a concept (construct) for analysis, the variables (indicators) selected to measure such concept must be both accurate and consistent (Hair *et al.*, 2019). Accuracy is associated with the term validity, while consistency is associated with the term reliability (Hair *et al.*, 2019). Thus, to ensure accuracy and consistency in the analysis, the assessment of the measurement model is designed to evaluate the assumptions pertaining to the validity and reliability of the structural model as identified in the SmartPLS software (SAGE Research Methods Datasets, 2019). In the measurement model using a PLS analysis, the validation guidelines of Hair *et al.* (2017) are used to examine the reliability and validity of the constructs along with their corresponding items.

To test the internal consistency reliability, Cronbach's alpha and composite reliability or simply CR, were calculated using Equations 3.1 and 3.2, respectively. Cronbach's alpha measures the internal consistency or reliability of a construct by considering how near and proximate the items are as a group (SAGE Research Methods Datasets, 2019; Taber, 2018). As opposed to Cronbach's Alpha, CR considers varying factor loadings. It is computed by getting the sum of all true score variables and covariances in the composite of indicator variables pertaining to the construct and dividing this sum by the total variance in the composite (SAGE Research Methods Datasets, 2019).

$$\alpha = \frac{K \cdot \bar{r}}{[1 + (K - 1) \cdot \bar{r}]} \qquad \textit{Equation 3.1}$$

where: K = the number of indicators for a given construct; $\bar{r}$= the average non redundant indicator correlation coefficient (*i.e.* the mean of the lower or upper triangular correlation matrix)

$$CR = \frac{(\sum_{k=1}^{K} l_k)^2}{(\sum_{k=1}^{K} l_k)^2 + \sum_{k=1}^{K} var(e_k)}$$ *Equation 3.2*

where: l_k = standardized outer loading of the indicator variable k of a specific construct measured with K indicators; e_k = measurement error of indicator variable k; $var(e_k)$ = the variance of the measurement error $(1 - l^2{}_k)$

The calculated values of internal consistency and convergent validity for participants based on nationality, prior IVR experience, and length of work experience are shown in Table 3.3, Table 3.4, and Table 3.5, respectively. As observed in Tables 3.3-3.5, Cronbach's alpha and CR values for all factors were above the minimum cut-off 0.6 (Hair *et al.*, 2017), which indicates that the constructs have strong internal consistency reliability for each considered sub-population. In other words, the indicators (question items) considered in this study provide an acceptable degree of consistency with respect to the intended construct (Taber, 2018).

Table 3.3. Internal consistency reliability and convergent validity analysis for participants based on nationality.

Constructs	**Items**	**Factor Loading**	**Cronbach's Alpha**	**CR[a]**	**AVE[b]**
Performance Expectancy	PE_1 PE_2 PE_3	0.907 (0.915) 0.918 (0.915) 0.899 (0.917)	0.893 (0.903)	0.934 (0.939)	0.824 (0.838)
Effort Expectancy	EE_1 EE_2 EE_3*	0.892 (0.898) 0.879 (0.862)	0.725 (0.711)	0.879 (0.873)	0.784 (0.775)
Social Influence	SI_1 SI_2 SI_3	0.765 (0.822) 0.753 (0.856) 0.823 (0.850)	0.680 (0.796)	0.824 (0.880)	0.610 (0.710)
Hedonic Motivation	HM_1 HM_2 HM_3	0.866 (0.798) 0.824 (0.750) 0.859 (0.766)	0.809 (0.682)	0.886 (0.815)	0.722 (0.595)
Behavioural Intention	BI_1 BI_2 BI_3	0.923 (0.935) 0.934 (0.945) 0.890 (0.842)	0.904 (0.894)	0.940 (0.934)	0.839 (0.826)

Note: * - Removed due to the lack of outer loading reliability (< 0.7)
[a] - Composite Reliability
[b] - Average Variance Extracted
Numbers in bracket - Values for Eastern countries group
Numbers not in bracket - Values for Western countries group

Table 3.4. Internal consistency reliability and convergent validity analysis for participants based on prior IVR experience.

Constructs	Items	Factor Loading	Cronbach's Alpha	CR[a]	AVE[b]
Performance Expectancy	PE_1	0.893 (0.917)	0.880 (0.910)	0.926 (0.943)	0.806 (0.847)
	PE_2	0.909 (0.926)			
	PE_3	0.892 (0.919)			
Effort Expectancy	EE_1	0.880 (0.906)	0.625 (0.757)	0.841 (0.892)	0.726 (0.804)
	EE_2	0.823 (0.887)			
	EE_3*				
Social Influence	SI_1	0.723 (0.774)	0.634 (0.743)	0.803 (0.853)	0.577 (0.660)
	SI_2	0.711 (0.825)			
	SI_3	0.839 (0.836)			
Hedonic Motivation	HM_1	0.830 (0.860)	0.672 (0.804)	0.816 (0.882)	0.597 (0.714)
	HM_2	0.722 (0.825)			
	HM_3	0.763 (0.849)			
Behavioural Intention	BI_1	0.918 (0.928)	0.878 (0.912)	0.925 (0.945)	0.804 (0.851)
	BI_2	0.904 (0.953)			
	BI_3	0.866 (0.886)			

Note: * - Removed due to the lack of outer loading reliability (< 0.7)
[a] - Composite Reliability
[b] - Average Variance Extracted
Numbers in bracket - Values for without prior experience to IVR group
Numbers not in bracket - Values for with prior experience to IVR group

Table 3.5. Internal consistency reliability and convergent validity analysis for participants based on length of work experience.

Constructs	Items	Factor Loading	Cronbach's Alpha	CR[a]	AVE[b]
Performance Expectancy	PE_1	0.913 (0.909)	0.913 (0.892)	0.945 (0.933)	0.851 (0.822)
	PE_2	0.933 (0.912)			
	PE_3	0.922 (0.899)			
Effort Expectancy	EE_1	0.886 (0.898)	0.712 (0.723)	0.874 (0.878)	0.776 (0.782)
	EE_2	0.876 (0.871)			
	EE_3*				
Social Influence	SI_1	0.767 (0.773)	0.710 (0.712)	0.837 (0.839)	0.632 (0.635)
	SI_2	0.807 (0.775)			
	SI_3	0.811 (0.840)			
Hedonic Motivation	HM_1	0.807 (0.865)	0.718 (0.795)	0.833 (0.879)	0.624 (0.708)
	HM_2	0.795 (0.803)			
	HM_3	0.767 (0.854)			
Behavioural Intention	BI_1	0.917 (0.928)	0.876 (0.912)	0.924 (0.945)	0.802 (0.851)
	BI_2	0.916 (0.946)			
	BI_3	0.852 (0.893)			

Note: * - Removed due to the lack of outer loading reliability (< 0.7)
[a] - Composite Reliability
[b] - Average Variance Extracted
Numbers in bracket - Values for more than 5 year work experience group

Numbers not in bracket - Values for less than 5 year work experience group

In order to test the construct validity, the convergent validity and the discriminant validity were calculated. Convergent validity evaluates the proximity or relation of items that measures the same construct while the discriminant validity determines whether a construct, which in theory must be unrelated from other constructs in the model, is indeed distinct and unrelated (Hair *et al.*, 2019). According to Hair *et al.* (2017), the calculated value for the average variance extracted or simply AVE (Equation 3.3) and the factor loading (*i.e.* coefficient of a question item in relation to specific construct) should be greater than 0.5 and 0.708, respectively. As shown in Tables 3.3-3.5, all constructs in every subpopulation considered had an average variance extracted (AVE) and factor loading values higher than the minimum cut-off. These results indicate that the indicators (questionnaire items) of a specific construct (PE, EE, SI, HM, or BI) positively converge, share, and measure the same construct (Hair *et al.* 2017).

$$AVE = \frac{(\sum_{k=1}^{K} l^2{}_k)}{K} \qquad \textit{Equation 3.3}$$

Although most of the researchers use the Fornell-Larcker criterion to test discriminant validity in PLS-SEM studies, there are studies which question the correctness of this criterion (Hair *et al.*, 2019; Henseler *et al.*, 2015). For instance, the simulation study of Henseler *et al.* (2015) showed that the Fornell-Larcker criterion was unable to perform and detect the lack of discriminant validity in PLS-SEM. This observation was particularly evident when the difference between the indicator loadings on a construct were small. To minimize this error, Henseler *et al.* (2015) suggested that the Heterotrait-Monotrait ratio (HTMT) should be used to evaluate the correlations among the measures of potentially overlapping constructs. Thus, for this study, the Heterotrait-Monotrait ratio (HTMT) was calculated to evaluate discriminant validity (Equation 3.4). The calculated values of discriminant validity for participants based on nationality, prior IVR experience, and the length of work experience is shown in Table 3.6, Table 3.7, and Table 3.8, respectively. As shown in Tables 3.6-3.8, the calculated confidence interval of the HTMT statistics was lower than the threshold value of 0.9 for all combination of constructs (Gold *et al.*, 2001). Thus, the results obtained indicate adequate convergent and discriminant validities for each considered sub-population.

$$HTMT_{ij} = \frac{1}{K_i K_j} \sum_{g=1}^{K_i} \sum_{h=1}^{K_j} r_{ig,jh} \div \left(\frac{2}{K_i(K_i-1)} \cdot \sum_{g=1}^{K_i-1} \sum_{h=g+1}^{K_i} r_{ig,ih} \cdot \frac{2}{K_i(K_i-1)} \cdot \sum_{g=1}^{K_j-1} \sum_{h=g+1}^{K_j} r_{jg,jh} \right)^{\frac{1}{2}} \quad \textit{Equation 3.4}$$

$$c = cor(Y^{(1)}, Y^{(2)})$$

where: c = correlation between the composite scores $Y^{(1)}$ and $Y^{(2)}$

Table 3.6. Discriminant validity analysis using Heterotrait-Monotrait (HTMT) ratio for participants based on nationality.

	PE	EE	SI	HM	BI
PE					
EE	0.813 (0.767)				
SI	0.568 (0.470)	0.595 (0.555)			
HM	0.781 (0.571)	0.804 (0.691)	0.523 (0.148)		
BI	0.867 (0.771)	0.887 (0.832)	0.590 (0.518)	0.844 (0.646)	

Note: The numbers indicate the pairwise correlations between variables
PE = Performance Expectancy; EE = Effort Expectancy; SI = Social Influence; HM = Hedonic Motivation; BI = Behavioural Intention
Numbers in bracket - Values for Eastern countries group
Numbers not in bracket - Values for Western countries group

Table 3.7. Discriminant validity analysis using Heterotrait-Monotrait (HTMT) ratio for participants based on prior IVR experience.

	PE	EE	SI	HM	BI
PE					
EE	0.746 (0.847)				
SI	0.452 (0.561)	0.426 (0.641)			
HM	0.553 (0.794)	0.756 (0.767)	0.306 (0.473)		
BI	0.785 (0.882)	0.830 (0.889)	0.466 (0.608)	0.687 (0.832)	

Note: The numbers indicate the pairwise correlations between variables
PE = Performance Expectancy; EE = Effort Expectancy; SI = Social Influence; HM = Hedonic Motivation; BI = Behavioural Intention
Numbers in bracket - Values for without prior experience to IVR group
Numbers not in bracket - Values for with prior experience to IVR group

Table 3.8. Discriminant validity analysis using Heterotrait-Monotrait (HTMT) ratio for participants based on length of work experience.

	PE	EE	SI	HM	BI
PE					
EE	0.819 (0.808)				
SI	0.538 (0.528)	0.552 (0.597)			
HM	0.636 (0.766)	0.799 (0.764)	0.231 (0.491)		
BI	0.824 (0.861)	0.895 (0.871)	0.542 (0.581)	0.779 (0.805)	

Note: The numbers indicate the pairwise correlations between variables

PE = Performance Expectancy; EE = Effort Expectancy; SI = Social Influence; HM = Hedonic Motivation; BI = Behavioural Intention
Numbers in bracket - Values for more than 5 year work experience group
Numbers not in bracket - Values for less than 5 year work experience group

3.3.3. Assessment of measurement invariance

Before performing the multi-group analysis (PLS-MGA) to determine the potential similarities and/or differences between path coefficients of the subpopulations considered, measurement invariance should be tested (Henseler *et al.*, 2016). Measurement invariance is required to ensure that a given measure is interpreted in a conceptually similar matter across a specified population (Horn & Mcardle, 1992). In PLS-SEM, measurement invariance can be tested using the measurement invariance of composites (MICOM) procedure which includes configural invariance, compositional invariance, and equality of composite mean values and variances (Henseler *et al.*, 2016). According to Henseler *et al.*, (2016), completing these three steps will give a full measurement invariance (*i.e.* pooling data of different groups), but establishing the first two steps is sufficient to conduct PLS-MGA.

The assessment of configural invariance involves the evaluation of the measurement models for all groups to check if the same number of indicators and the same variance-based model estimation were used and if all the indicators were treated equally across the specified groups (Henseler *et al.*, 2016). As the analysis and assessment of the measurement models (reliability and validity) for all groups was completed in the previous sub-section, configural invariance was established.

To ensure the homogeneity of the composite scores across the considered subpopulations, compositional invariance was examined using a permutation analysis with 5000 resamples through SmartPLS 3 software (Henseler *et al.*, 2016; Ringle *et al.*, 2015). The calculated values of MICOM for participants based on nationality, prior IVR experience, and length of work experience are shown in Table 3.9, Table 3.10, and Table 3.11, respectively. As shown in Tables 3.9-3.11, all of the values of *c* (compositional invariance correlation) were close to 1 and fell within the 95% confidence interval. Hence, the compositional invariance was established across all the subpopulation groups. Subsequently, upon the establishment of both configural and compositional invariance, partial measurement invariance, which is the minimum required to conduct the PLS-MGA was achieved.

Table 3.9. Assessment of measurement invariance of composite models (MICOM) test for participants based on nationality.

Construct	Step 1

	Configural Invariance		
PE	**Yes**		
EE	**Yes**		
SI	**Yes**		
HM	**Yes**		
BI	**Yes**		
Construct	**Step 2**		
	Correlation c	95% confidence interval	Compositional invariance
PE	1.000	[0.999, 1.000]	**Yes**
EE	1.000	[0.997, 1.000]	**Yes**
SI	0.998	[0.984, 1.000]	**Yes**
HM	0.997	[0.992, 1.000]	**Yes**
BI	1.000	[0.999, 1.000]	**Yes**

Note: PE = Performance Expectancy; EE = Effort Expectancy; SI = Social Influence; HM = Hedonic Motivation; BI = Behavioural Intention

Table 3.10. Assessment of measurement invariance of composite models (MICOM) test for participants based on prior IVR experience.

Construct	**Step 1**		
	Configural Invariance		
PE	**Yes**		
EE	**Yes**		
SI	**Yes**		
HM	**Yes**		
BI	**Yes**		
Construct	**Step 2**		
	Correlation c	95% confidence interval	Compositional invariance
PE	1.000	[0.999, 1.000]	**Yes**
EE	1.000	[0.997, 1.000]	**Yes**
SI	0.997	[0.983, 1.000]	**Yes**
HM	0.999	[0.992, 1.000]	**Yes**
BI	1.000	[0.999, 1.000]	**Yes**

Note: PE = Performance Expectancy; EE = Effort Expectancy; SI = Social Influence; HM = Hedonic Motivation; BI = Behavioural Intention

Table 3.11. Assessment of measurement invariance of composite models (MICOM) test for participants based on length of work experience.

Construct	**Step 1**		
	Configural Invariance		
PE	**Yes**		
EE	**Yes**		
SI	**Yes**		
HM	**Yes**		
BI	**Yes**		
Construct	**Step 2**		
	Correlation c	95% confidence interval	Compositional invariance
PE	1.000	[0.999, 1.000]	**Yes**
EE	1.000	[0.997, 1.000]	**Yes**

SI	0.998	[0.985, 1.000]	**Yes**
HM	0.996	[0.993, 1.000]	**Yes**
BI	1.000	[0.999, 1.000]	**Yes**

Note: PE = Performance Expectancy; EE = Effort Expectancy; SI = Social Influence; HM = Hedonic Motivation; BI = Behavioural Intention

3.3.4. Assessment of the structural model and PLS-MGA results

After establishing the minimum requirement to perform PLS-MGA, the structural model was evaluated for every subpopulation based on the collinearity assessment, the coefficient of determination (R^2), and the path coefficient significance (β). Before calculating R^2 and β, it is important to check first if there are no issues connected with multi-collinearity (Hair *et al.*, 2017). To do this, full collinearity variance inflation factors (VIFs) are evaluated. As shown in Tables 3.12, as the obtained VIF values for PE, EE, SI, and HM were significantly below the threshold value of 3, multi-collinearity issues were not a concern.

Table 3.12. Results of full collinearity test for each subpopulation.

Construct	**Variance Inflation Factors (VIF)**					
	Western	Eastern	With Prior IVR experience	Without Prior IVR experience	< 5 year work experience	> 5 year work experience
PE	2.283	1.785	1.632	2.634	2.043	2.237
EE	2.016	1.973	1.678	2.257	2.219	1.976
SI	1.298	1.299	1.143	1.360	1.297	1.290
HM	2.063	1.488	1.433	2.068	1.748	1.918

Note: PE = Performance Expectancy; EE = Effort Expectancy; SI = Social Influence; HM = Hedonic Motivation; BI = Behavioural Intention

The structural model that specifies the correlations between the constructs for each subpopulation was evaluated by investigating the path significance using a bias-correlated and accelerated (BCA) bootstrapping without sign change re-sampling technique based on 5000 sub-sample (Hair *et al.*, 2017; Ringle *et al.*, 2015). BCA bootstrapping was used to handle the issue of peaked and skewed distribution by adjusting the confidence intervals for skewness (Efron, 1987). The path coefficients and the extent of influence on the structural equation model for every subpopulation are shown in Figures 3.4-3.6.

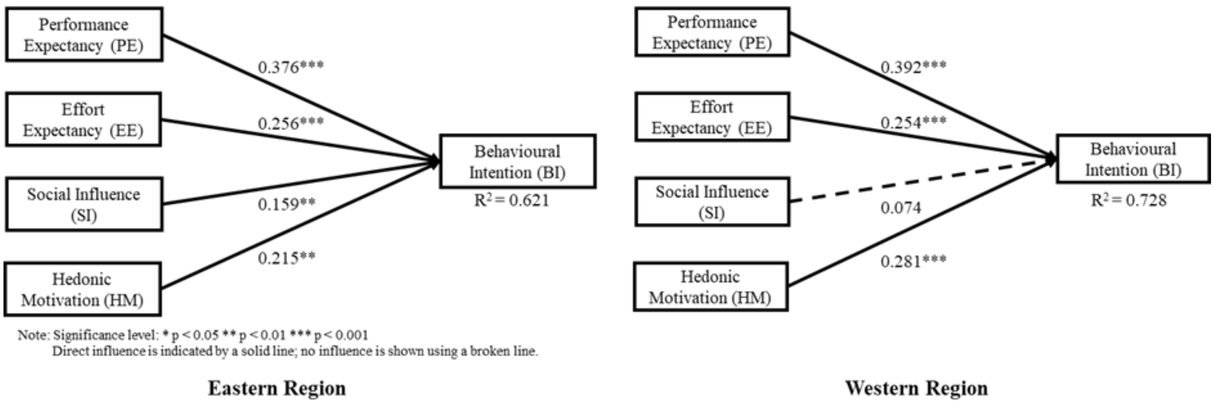


Figure 3.4. Structural equation model of the employees' perception on IVR games in training based on nationality. Significance Level: *p< 0.05**p<0.01***p<0.001. Direct influence is indicated by a solid line; no influence is shown using a broken line.

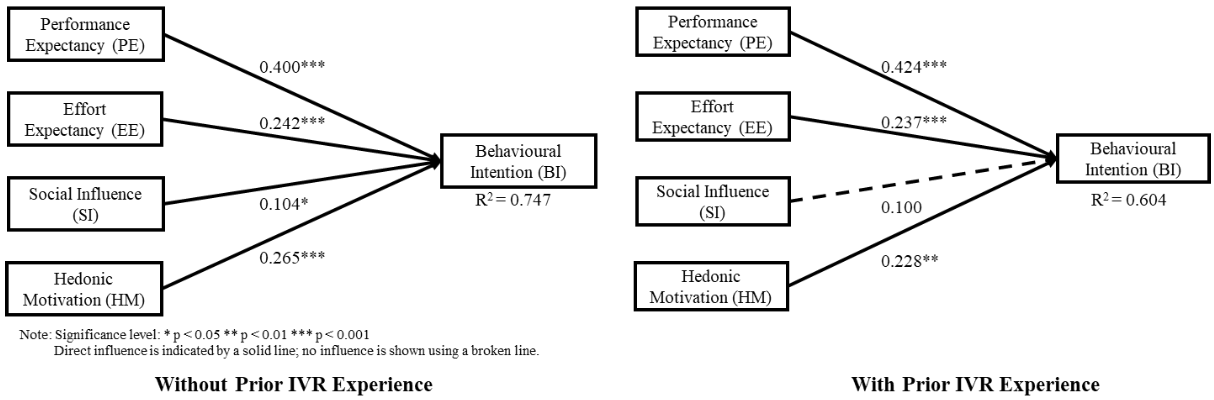


Figure 3.5. Structural equation model of the employees' perception on IVR games in training based on prior IVR experience. Significance Level: *p< 0.05**p<0.01***p<0.001. Direct influence is indicated by a solid line; no influence is shown using a broken line.

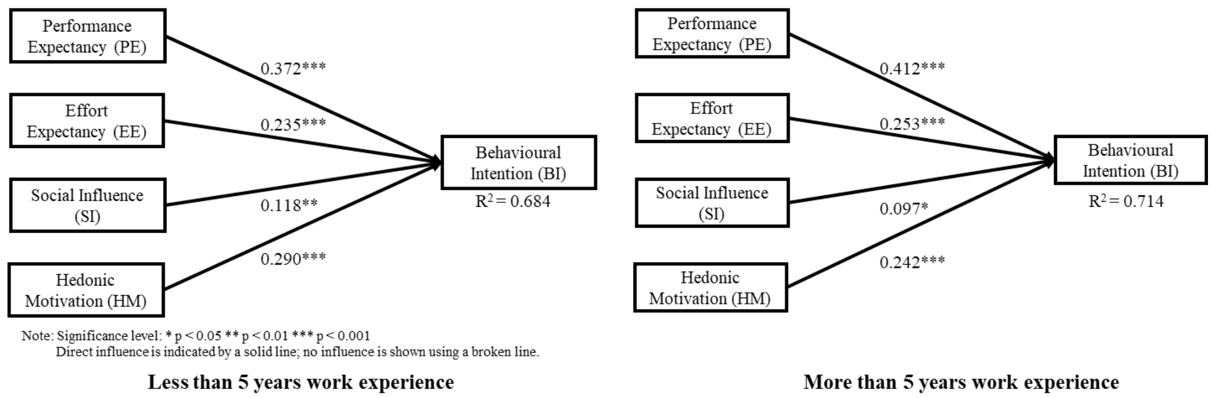


Figure 3.6. Structural equation model of the employees' perception on IVR games in training based on length of work experience. Significance Level: *p< 0.05**p<0.01***p<0.001. Direct influence is indicated by a solid line; no influence is shown using a broken line.

As shown in figures 3.4-3.6, the variance explained by the PE, EE, SI, and HM constructs for behavioural intention to adopt IVR was 0.728, 0.621, 0.604, 0.747, 0.684, and 0.714 for the groups from the western region, eastern region, with prior IVR experience, without prior IVR experience, less than 5 years of work experience, and more than 5 years of work experience, respectively. Since R^2 value > 0.2 is considered acceptable in the behavioural study, all models possess adequate capacity to explain BI to adopt IVR (Hair *et al.*, 2017).

Table 3.13 summarises the outcomes of the path coefficient values for each subpopulation. H1, which suggested significant relationships between PE and BI, was validated in all groups and the values were above 0.37. H2, which suggested significant relationships between EE and BI, was validated in all groups and the values were above 0.23. H3, which suggested significant relationships between SI and BI, was not validated in the western group nor in the prior IVR experience group, but was validated in the other groups. H4, which suggested significant relationships between HM and BI, was validated in all groups and the values were above 0.21.

Table 3.13. Outcomes of the Structural equation model multi-group analysis.

Relationship	Based on nationality			
	Eastern	Western	∣Diff∣	Henseler's MGA *p*-value
H1: PE → BI	0.376***	0.392***	0.016	0.566
H2: EE → BI	0.256***	0.254***	0.002	0.482
H3: SI → BI	0.159**	0.074	0.084	0.108
H4: HM → BI	0.215**	0.281***	0.066	0.795
Relationship	Based on prior experience to IVR			
	Without IVR Experience	With IVR Experience	∣Diff∣	Henseler's MGA *p*-value
H1: PE → BI	0.400***	0.424***	0.024	0.605
H2: EE → BI	0.242***	0.237***	0.005	0.477
H3: SI → BI	0.104*	0.100	0.004	0.474
H4: HM → BI	0.265***	0.228**	0.037	0.319
Relationship	Based on length of work experience			
	< 5 years	> 5 years	∣Diff∣	Henseler's MGA *p*-value
H1: PE → BI	0.372***	0.412***	0.041	0.685
H2: EE → BI	0.235***	0.253***	0.018	0.587
H3: SI → BI	0.118**	0.097*	0.021	0.370
H4: HM → BI	0.290***	0.242***	0.049	0.251

Note: Significance level of path coefficient: * p < 0.05, ** p < 0.01, *** p < 0.001.
Diff. = Path Coefficient Differences

PE = Performance Expectancy; EE = Effort Expectancy; SI = Social Influence; HM = Hedonic Motivation; BI = Behavioural Intention

Having evaluated the measurement and structural model, Henseler's MGA (PLS-MGA), a non-parametric test, was used to assess the similarities and differences of path coefficients between the groups. In this method, if the MGA *p*-value is lower than 0.05 or greater than 0.95, there is a 5% level significant difference between specific path coefficients between the two subpopulations. The outcome of the PLS-MGA *p*-values in Table 3.13 showed that there were no significant group differences between any of the subpopulation groups (*e.g.* based on nationality, prior IVR experience, and length of work experience). Therefore, H5, H6, and H7 were not accepted.

3.4. Discussion

The current study applied a group-based approach to examine the perception of chemical operators and employees towards IVR adoption intentions. This was done by using a modified version of UTAUT 2 that includes PE, EE, SI, and HM constructs. Upon comparison of the groups of chemical operators and employees based on nationality, prior IVR experience, and work experience using PLS-MGA, there were several similarities and differences in the relationships investigated in the current study.

3.4.1. Theoretical implications

This study provides meaningful insights for the current literature on IVR adoption based on the UTAUT 2 model. Structural equation models developed for this purpose indicated that PE, EE, and HM have a significant impact on BI for all subpopulations (western, eastern, with and without prior IVR experience, less than and more than 5 years of work experience) while the construct SI has a significant impact on BI for subpopulations such as eastern, without prior IVR experience, as indicated by the parameters such as path coefficients and p-values. The limitation of data-based modelling approaches is their inability to relate the independent variables (in this case PE, EE, SI, and HM) to other independent variables as this is beyond the scope of the UTAUT2 model. However, Sitar-Tăut (2021) pointed out that there was a causal connection between the independent variables PE and HM on mobile learning using their proposed acceptance framework. This study does not claim such causal connection, but the reliability and validity testing carried out within this research ensures the independence of the latent variables. Moreover, the explanatory power, determined by the R^2 values computed for each group-specific structural equation models, all have values greater than 60.4%. These R^2 values are greater than the 40% recommended value for behavioural intention (Zhou *et al.*, 2021). Thus, future practitioners such as instructors and researchers can use this methodology based on the modified UTAUT 2 model on IVR adoption intention to test its reliability and validity in other settings.

Concerning hypothesis testing, the empirical results for all subpopulations (western, eastern, with and without prior IVR experience, less than and more than 5 years of work experience) showed that performance expectancy significantly influences the IVR adoption intention in chemical industries. Moreover, among the four key factors, PE was the strongest factor influencing BI to adopt IVR for all subpopulations (Figures 2-4). This result is also consistent with previous studies that confirmed the significant influence of PE on VR acceptance in a sport context (Kunz & Santomier, 2019) and on e-mail acceptance (Mao & Palvia, 2008). This indicates that when users know that the given technology provides better job performance (*e.g.* enhancing effectiveness, increasing productivity, improving productivity, *etc.*), they will become more motivated to adopt the system. As Hsu and Lin (2008) mentioned that PE is a critical factor in work-related environment and given that the respondents considered in this study were chemical operators and employees, this suggests that regardless of the group, respondents perceive that using IVR technology in chemical industry training would increase their job performance.

In addition, effort expectancy was found to significantly influence the IVR adoption intention for all subpopulations. The results show that regardless of the group, respondents are driven to adopt IVR if they perceive the IVR experience as easy and simple. This outcome is also consistent with the findings on mobile internet acceptance (Venkatesh *et al.*, 2012) and online virtual tour-guiding platform acceptance (Chiao *et al.*, 2018). This suggests that when users know that a given platform is user-friendly, it boosts their confidence which leads to adopting and supporting that system. However, the rank order for EE is different per subpopulation. For instance, subpopulations such as eastern group, with prior IVR experience group, and more than 5 years of work experience group show that EE is the second strongest factor influencing BI to adopt IVR. According to Ramayah *et al.* (2005), EE is considered to have more influence on BI for less experienced users but, in this study, this was not verified. This may be because VR is still not that accessible compared to other technologies, such as desktop computers or smartphones. It is also possible that the conflicting result is due to the fact that the respondents with prior IVR experience probably still expected it not to be easy and to take quite some time for them to master the controls in the IVR environment. Moreover, compared to less than 5 years of work experience, respondents who have more than 5 years of experience are more hesitant to use the IVR training. For instance, Elgohary and Abdelazyz (2020) reported that the people who have more years of work experience have a higher tendency to be hesitant, less confident, and comfortable in using new technology as opposed to the platforms they were already accustomed to. Thus, it is understandable that for these groups, the amount of effort needed by the user to put into using IVR-based training (EE) is a more important factor to consider for IVR adoption intention than SI or HM. Furthermore, compared to those respondents who come from the western region, respondents come from eastern region are more reluctant to use IVR-based training. According to Cipresso *et al.* (2018), European countries such United Kingdom, Germany, Italy, Spain, among others were very involved in VR research. Given that several chemical institutions in the abovementioned European countries have actively participated in the development of new IVR-based training that can be used in their respective needs, their employees have greater chance to access these new technologies. This reason could explain the lower path coefficient effort expectancy value for respondents coming from western group compared to eastern group.

Hedonic motivation also significantly influenced the IVR adoption intention for all subpopulations. The results show that, regardless of the group, respondents are driven to adopt IVR if they perceive the IVR experience as fun and entertaining. This outcome is also consistent with previous studies on the IVR adoption intentions for an e-scooter service (Huang, 2020), and the acceptance of social telepresence robots (J. Han & Conti, 2020). Those studies indicated that close interaction between the user and the given technology can provide perceived fun-filled user experience. However, the rank order for HM is different in each subpopulation. For instance, subpopulations such as western group, without prior IVR experience group, and less than 5 years of work experience group, all show that HM is the second strongest factor influencing BI to adopt IVR. As mentioned by Venkatesh *et al.* (2012), as experience intensifies, the attractiveness of the novelty and innovativeness will reduce which in return affects the HM. Given this argument, it is possible that respondents who do not have prior IVR experience are more eager to experience the IVR technology than the respondents with prior IVR experience. This is also true for respondents with less than 5 years of work experience as well as for the western group. This may be because respondents coming from these groups are more willing to go out of their comfort zones and are more open to trying new and innovative approaches, such as IVR technology. Given these reasons, for the abovementioned groups, even though IVR technology will be mainly used for learning procedural know-how, HM is a more important factor than SI or EE, when considering IVR adoption intention.

Although the rank order is last, subpopulations such as eastern, without prior IVR experience, regardless of the length of work experience, indicated that social influence significantly impacts IVR adoption intention in chemical industries. This outcome is also consistent with previous studies that confirmed the significant influence of SI on e-governance of users (Alraja, 2016), and also on health IT (Bozan *et al.*, 2016) adoption intentions. Similar to the *"bandwagon effect"*, people tend to adopt new technology if it works favourably for their respected peers and/or supervisors (Tsai *et al.*, 2013). This effect is especially true in a situation where the implementation of a new technology is still in its initial stage as reported by Alraja, (2016). Thus, this suggests that for these groups, peer influence is still considered to be an important factor in determining what people adopt since the implementation of IVR is in its initial stage. However, subpopulations such as western, and with prior IVR experience groups do not support this hypothesis. Researchers, such as T. Teo and Noyes (2014), also reported that SI was found not to be significant on BI to use technology among younger pre-service teachers as they were digital natives and tended to choose for themselves whether to use the given technology or not. For the two groups mentioned above, it is possible that people within these groups were already aware of the existence of IVR technology (*i.e.* digital natives). This suggests that they really do not need to be influenced by their peers or supervisors as they know the capabilities of IVR technology. Thus, it is important to consider the subpopulation as the construct of SI changes over time.

The modified UTAUT2 is helpful in exploring more factors that influence the intentions of chemical operators and employees to adopt IVR in a different setting. Through using this model, the relationships between BI and PE, EE, SI, and HM were explored. The results of the analysis and the Henseler's MGA analysis revealed that there was no significant difference in the models between the effect of PE, EE, SI, and HM on BI to adopt IVR in a group of chemical operators and employees based on nationality, prior IVR experience, or length of work experience. Nevertheless, it is still necessary to take into account these socio-demographic factors as there are definite group differences in terms of the ranking order of each construct for the IVR adoption intentions among each subpopulation. Incorporating PLS-SEM and MGA methods is beneficial since these methods are not limited to analysing IVR adoption behaviour of the population sample, but they are also useful in determining group differences (Matthews, 2017; Ramírez-Correa *et al.*, 2015).

3.4.2. Practical implications

In terms of the practical implications, this study will be able to inform the chemical industry policymakers/decision-makers in a number of ways. If the institutions decide to create an IVR-based training, they can consider the ranking order of each construct to design appropriate training strategies in the IVR environment to satisfy the needs of the users. Through this, managerial and training staff, and corporate policymakers will have a clearer view on what should be implemented, allowing them to decide whether to emphasize game elements, easier controls, more procedural aspects, *etc*. Purchase the IVR system from other industries can also be guided through the use of the results of this model, as a basis to choose the most suitable IVR system and to make a fair judgement regarding the IVR system specifications that lead to a more effective delivery of the training programme. As the COVID-19 pandemic continues to develop through multiple waves around the world, online courses, as well as training, are becoming the new normal. Thus it is expected that IVR will play a significant role in delivering professional development and health and safety training. The current study recommends that it is important to consider constructs such as PE, EE, SI, and HM as key factors in determining the adoption rate of IVR technology.

3.5. Conclusions

This study aimed to compare the factors influencing the adoption intention of IVR using the modified version of UTAUT 2 model. The IVR adoption intention of predefined groups of chemical operators and employees was analysed using PLS-SEM and multi-group analysis (MGA) with SmartPLS 3.0. Although the results of PLS-MGA did not show statistically significant differences between the predefined groups of respondents, the MGA approach is effective in understanding the intentions of multiple groups. Through identifying the ranking order of the constructs considered in the UTAUT model across group-specific results, chemical industry policymakers can utilise this information in formulating suitable strategies on possible ways to implement IVR-based technology from the measured groups (*e.g.* whether to emphasize game elements, easier controls, more procedural aspects, *etc.*).

3.6. Note

Part of this chapter has been published in the **Virtual Reality Journal**: Toyoda, R., Russo Abegão, F., Gill, S., & Glassey, J. (2021). Drivers of immersive virtual reality adoption intention: a multi-group analysis in chemical industry settings. Virtual Reality. https://doi.org/10.1007/s10055-021-00586-3

Chapter 4. A framework for evaluating user expertise in immersive learning environments

Regular and relevant training is helpful in ensuring that employees within an organization are provided with necessary skills and information for the efficient performance of their roles. One of the training methods now being used in industries is the immersive virtual reality (IVR) training. Given that the IVR-based training method enables provision of effective training in safe and controlled environments as perceived by the chemical operators and employees working at different chemical industries from chapter 3, there is a need from stakeholders for the development of assessment frameworks for training in IVR settings. This study provides a guideline on employing an unbiased and unobtrusive assessment framework for the evaluation of user expertise in immersive learning environments through applying Dreyfus model and evidence-centred design (ECD) and fuzzy comprehensive evaluation method. The developed assessment framework was implemented and tested in a case study which involves an IVR-training called ‘Operate Your Own Reactor’ where 19 chemical operator apprentices from a chemical training company based in Belgium were trained to perform tasks for the production of *n*-butyllithium in a virtual chemical plant. The assessment results provided by the IVR system automatically display a detailed performance level of the participants and corresponding feedback per task. This information can then be used by the apprentices to identify their respective strengths and weaknesses. Moreover, this assessment framework can assist trainers in providing more constructive and personalised feedback for trainees to ensure their continual growth. Lastly, the relevant and timely information about the performance of apprentices in IVR-based training, provided through this assessment framework, can be used by the administrators for strategic decision-making in terms of personnel management and development. In summary, this study creates an assessment system for a relatively new technology, the IVR, for chemical reactor operation training, through the application and harmonisation of proven and validated methodologies.

4.1. Introduction

The contributions of experts are necessary in advancing their respective fields. However, before a person becomes an expert, they must undergo an often extensive and lengthy process of knowledge and skill development. Achievement of expertise in a specific field, although difficult, may be significantly aided by creating a systematic, effective, and well-designed training programs It is therefore essential that the advancement of a novice towards an expert can be measured in definite levels. To assess this progression, researchers and practitioners have consistently used the Dreyfus model (Dreyfus, 2004; Dreyfus & Dreyfus, 1987) to describe the gradual process of levelling up towards expertise. The Dreyfus model provides a five-level hierarchy consisting of a novice, advanced-beginner, competent, proficient, and expert, wherein each level is characterized in terms of their traits and capacity in solving a specific problem. The incorporation of the Dreyfus model in an assessment framework will provide a concrete backbone for an effective training system deliberately designed to efficiently develop experts.

The outcomes of the training are greatly affected through the chosen methodologies for training design, delivery, and implementation (Salas *et al.*, 2012). In most cases, trainees tend to retain information better if they experience both talking and doing, rather than just receiving a long-written protocol to read and follow (Dholakiya *et al.*, 2019). As a result, several institutions are now employing new technologies, such as the immersive virtual reality (IVR) technology for training (Garcia Fracaro *et al.*, 2021; Isleyen & Duzgun, 2019). In contrast to the traditional training methods, IVR training simulates actual situations in a safe and realistic environment for the acquisition of knowledge and skills (Fällman *et al.*, 1999). IVR technology which incorporates games and simulations, is also shown to enhance problem-solving and decision-making skills (Voorhis & Paris, 2019).

However, several difficulties arise when developers of such interactive immersive environments attempt to classify and measure the competencies acquired. One of the difficulties is related to the method of training assessment. For instance, the assessment of trainees through paper-and-pencil tests is limited to measuring declarative knowledge acquisition rather than addressing the development of 21st century skills such as complex problem solving and communication, essential to succeeding in professional life (Garcia Fracaro *et al.*, 2021; Loh, 2011; Shute & Emihovich, 2018). Moreover, assessment made by the experts is time consuming and might involve their subjective opinion which might affect the assessment outcomes (Garcia Fracaro *et al.*, 2021). Aside from the fact that assessment needs to be carried out in an unbiased and unobtrusive way, information generated through the actions and behaviours of the users within the system should be utilised in the assessment in these kinds of virtual environments (Loh, 2011; Shute *et al.*, 2016).

The result of an effective assessment allows the stakeholders to utilise the information in productive ways (Clarke-Midura & Dede, 2010). Thus, several researchers have been working on how to design and develop performance-based assessments for hard-to-measure and complex constructs, such as leadership and teamwork traits, in immersive environments.

To significantly improve assessment in IVR, evidence-centred design (ECD), which is an assessment design framework that integrates conceptual and analytical models for an automated assessment process, may be incorporated therein (Mislevy *et al.*, 2003; Shute *et al.*, 2016). Using this framework in IVR setting, real-time estimates and claims of competency levels of the learners possessing a range of knowledge and skills can be obtained from their in-game actions and behaviours in relation to the given tasks (Mislevy *et al.*, 2003). More importantly, ECD also narrows down and identifies the most relevant data among the voluminous data available and produced in the IVR in relation to the specific training objectives. This increases the operational efficiency of the IVR training platform by collecting, analysing, and archiving only the useful data for training and assessment purposes.

However, when the assessors use linguistic descriptors, such as excellent, very good, good, fair or poor, to evaluate an attribute (*e.g.* knowledge, skills, and/or attitude), it is usually ambiguous as such descriptors are not easily measured and cannot be expressed accurately using typical binary methods. In order to deal with these imprecise and uncertain data, a fuzzy comprehensive evaluation method is incorporated in this research. This method is based on fuzzy set theory, which deals with the determination and inclusion of possible 'true' values in a scale, developed by Zadeh (1965) and subsequently and consistently used by other researchers, such as Lasunon, (2019), Chen *et al.* (2015), and Sudhagar and Ganesan (2011). As an example, consider a situation where a performance is assessed as "good" if the total time to finish a certain task in VR environment in less than or equal to 10 minutes. If the binary method is used to evaluate performance, then the value of 10.1 minutes is not considered as "good". But, by using fuzzy comprehensive evaluation method the value of 10.1 minutes will be evaluated to some degree of "good". The degree of membership is directly connected to the closeness of the given value to the predetermined value of the term "good".

As the process of evaluating knowledge or performance skills in IVR-based environment can be challenging without an appropriate assessment structure, the objectives of this study are (1) to develop a framework of assessment guidelines for the evaluation of expertise in immersive learning environments through the application of appropriate methodology; and (2) to evaluate the utility of the developed assessment framework in IVR environment through realistic case study.

4.2. Expertise Development

In establishing the assessment framework, reference must be made to the level of knowledge and skills of experts of each respective field. Experts are skilled individuals who are essential in the orderly operations of highly complex industries and services, such as chemical (Nazir *et al.*, 2012), surgical (Ghaderi *et al.*, 2021) and dental (Lyon, 2014) fields. However, producing the desired number of skilled professionals in these industries requires a very long and costly training processes. Identifying and distinguishing the representative features, such as cognitive, technical, and other competencies, of these skilled professionals is important for the creation of effective training method that can maximise the ability to train more skilled professionals in a shorter period of time. Thus, it is critical to understand the acquisition and development of the desired knowledge, skills, and/or attitudes by the skilled professionals and novices in the workplace using well-established expertise development theories (Persky & Robinson, 2017).

Research on behavioural and cognitive differences between skilled individuals and novices began during the 1960s and it is a well-studied phenomenon especially in the field of psychology and training literature (Attri, 2019; Ericsson *et al.*, 2006). Since then, several studies on expertise development and skill acquisition theories that describe the path as well as the corresponding developmental activities from novice to skilled individual have been developed, implemented, and evaluated (Grenier & Kehrhahn, 2008).

Categorising users into only two extreme categories, novices and experts, leaves the developmental process unclear. The ambiguity in only using two categories makes it difficult to identify the breadth of knowledge and skills of each incremental step towards expertise. Thus a number of authors, such as Alexander (2003) and Nichols (2009), developed and used a multi-stage model of expertise development, including a middle stage(s), in educational and learning context (Alexander, 2004).

One model in particular, the Dreyfus model, has been reviewed, validated, and continuously used by many researchers in various fields such as clinical (Peña, 2010), nursing (Benner, 1982), and engineering (Albers *et al.*, 2012). As a widely-used expertise development framework, researchers, such as Farrell (2012), Hall-Ellis and Grealy (2013), and Hunt and Weintraub (2007), find that the Dreyfus model significantly and adequately captures the incremental and procedural advancement of a novice towards becoming an expert. In this model, one progresses through five stages of proficiency in skills acquisition: from novice to advanced beginner, competent, proficient, and finally to expert (Dreyfus, 2004; Dreyfus & Dreyfus, 1987). This five-stage approach provides a more intricate system to delineate generic skills to more specific learning experiences (Farrell, 2012). The Dreyfus model is based on the idea that the skill acquisition of an individual is a continuous process. As a learner passes each stage of qualitatively and/or quantitatively different insights of the skill and/or mode of decision making, there will be a gradual shift toward higher proficiency, for example becoming more intuitive or acquiring situational awareness. Table 4.1 summarises the characteristics of the different competency levels of an individual.

Table 4.1. Characteristics of different competency levels of an individual (Dreyfus, 2004; Dreyfus & Dreyfus, 1987).

Competency level	Characteristics
Novice	Novices can only perform their tasks by following a set of guidelines. Thus, they have little situational perception and behave with restricted judgment. When confronted with complicated tasks, they need close supervision.
Advanced Beginner	At this level, the advanced beginner has some appreciation of specific details about learning but still has limited situational perception. They are developing an ability to evaluate relevance of information but may still tend to rely on guidebooks. As such, they provide limited answers to unusual or complicated tasks since they do not always realise the possible results. They can perform regular sequence of tasks under indirect supervision, but still need it for complicated tasks.
Competent	Competent individuals start to see how actions may affect long-term results. They can organise and assess the circumstances to concentrate on important details. They perform deliberate planning, understand procedures by experts, and can adapt to new settings. They can perform complex analytical tasks and planning but still need supervision for non-routine complicated tasks.
Proficient	Individuals at this level are more accountable and confident and can deal with complicated scenarios holistically. They comprehend with clarity good and relevant information to resolve problems. At this stage, they are beginning to develop intuitive judgment and solve problems based on prior knowledge. Proficient individuals understand rules, theories, and alternatives more deeply.
Expert	Experts are better in decision-making through intuitive reasoning. They use analytical approaches in adapting to new situations, and they no longer rely on policies and procedures. They have a sense of responsibility for themselves, others, and the environment while envisioning the overall picture and possible alternatives.

The abovementioned developmental characteristics can be seen as the results of a successful transformation of four areas of mental functions, namely - components (*i.e.* the contextual features of a scenario that learner can recognise), perspective (*i.e.* the ability of the learner to choose the most significant trait of a scenario), decision (*i.e.* the ability of the learner to choose between analytical and intuitive reasoning on a given scenario), and commitment (*i.e.* the degree to which the learner feels accountable when understanding and making judgement), which are correlated to the transformations per skill level (Dreyfus, 2004; Dreyfus & Dreyfus, 1987).

Experts accomplish their tasks by exercising context-free and situational problem solving (Dreyfus, 2004; Dreyfus & Dreyfus, 1987). They are also more experienced, intuitive, and involved in their functions, possess wider perspective, as well as enhanced decision and commitment functions in the skill hierarchy (Dreyfus, 2004; Dreyfus & Dreyfus, 1987). As such, many institutions use the behaviours and characteristics of experts as foundation and reference in designing training instructions (Ericsson *et al.*, 2006). In designing the assessment framework using the experts as reference, the traits of experts are characterised and scaled into several components which may be measured and grouped into several levels. These levels are intended to be emulated by non-experts in order for them to progress towards expertise. For instance, in achieving superior working memory of learners, it is important to create scenarios which place more emphasis on identifying how the information correlates with broader concepts rather than only memorising specific facts or procedures. The variables to indicate the expertise level may be evaluated using qualitative methods, such as peer-evaluation, and quantitative methods, such as formulating criteria corresponding to number of mistakes within a certain time frame (Loh & Sheng, 2013; Unsworth, 2001).

Both qualitative and quantitative analysis allow tracking of improvement and/or worsening of the performance of the trainees through on the Dreyfus model. Since the success or failure of evaluating the levelling up process of the learner depends on gathering and selecting useful information from the given data, it is important to develop and incorporate a tool capable of extracting, organising, and quantifying actual performance data of the trainees in a specific training.

4.3. Evidence-centred design

The Dreyfus model provides guidance on what concepts trainees are expected to know and what practices trainees are expected to be able to demonstrate within a certain stage (Dreyfus, 2004). However, the model does not define detailed examples of the model-aligned curricula and assessments especially in IVR environment. Since every collected piece of information in the IVR environment (*e.g.* mistakes committed, total task duration, hints received, *etc.*) must be aligned with real-life learning and assessment, it is important to develop a valid performance-based assessment for hard-to-measure constructs (*e.g.* knowledge and skills). These constructs must accurately assess how trainees use complex competencies that are directly relevant to the skill acquisition model of the real world.

While there are models that relates collected data and information to relevant assessment constructs, such as task-centred approach, several authors, such as Snow *et al.* (2010), Cukurova *et al.* (2017), and Oliveri *et al.* (2019), prefer and used ECD as it allows the collection of relevant information from performance data (observable data) and relates these to theoretical (unobservable) constructs through evidentiary arguments (Mislevy *et al.*, 2003). This is important for accountability purposes of the stakeholders involved (Zieky, 2014). ECD also provides a blueprint for designing and redesigning assessment frameworks that may be utilised in diverse settings (Mislevy *et al.*, 2017). It also allows the gathering of various types of data and thus increases the reliability and validity of the assessment during the test design and development process (Zieky, 2014). Furthermore, ECD is flexible as it can be used for various purposes, for example formative, summative assessments, and for assessing various types of learner attributes (Snow *et al.*, 2010). In this context, the incorporation of the ECD in the assessment framework will allow an accurate method for measuring complex competencies in the IVR environment.

As ECD requires the development of an evidence model to assess and relate measurable attributes to the proficiency of the learners, it is necessary to create a detailed and robust coding system or scoring rubrics to understand the context of the tasks and avoid subjective judgments (Mislevy *et al.*, 2017). Further, ECD requires flexibility as to the rigid application of data collection structure and the probability of including other data which may be a potential source of relevant information. (Mislevy & Haertel, 2007; Mislevy & Riconscente, 2005). In general, ECD may be time-consuming and requires additional detailed steps. However, its benefits in providing accurate, valid, and relevant assessment for 21st century competencies far outweigh the apparent disadvantages (Zieky, 2014).

Primarily, ECD involves the collection of essential information about the subject that has a direct relation to the assessment of the skills or mastery of a person, or lack thereof. This will then require the conceptualisation of the assessment argument to link the gathered information to the skills being assessed. Afterwards, the conceptual assessment framework (CAF) shall be created to serve as blueprint for tasks, evaluation procedures, and measurement models (Mislevy & Haertel, 2007).

Some components of ECD framework must be given more importance than others depending on the training environment and its objectives (Mislevy *et al.*, 2003). In IVR environment, the simulation of a real-world scenario must have a deliberate design, in order to collect relevant evidence as to the trainee's competency. Thus, in this platform, the framework must primarily identify its goal and purpose in order to define the competencies necessary to evaluate whether or not such purpose was achieved. Corollary, the framework must establish the variables and their levels, measurable in the IVR environment, corresponding to the competencies.

To complete the assessment framework design in the IVR environment, it is necessary to review existing literature and interview field experts to analyse cognitive tasks *vis-à-vis* measurement of competencies. Through such activities, scenarios, involving various problems and situations trainees realistically deal with, will be specifically designed to elicit data and information. The data and information that are gathered in the training must be correlated, organised, and designed to measure and evaluate the trainees through their specific characteristics (competency model), what manifested through their statements and actions (evidentiary model), and how they acted in each task (task model).

4.4. Fuzzy comprehensive evaluation model

The aim of fuzzy comprehensive evaluation model is to provide further delineation between each of the competencies in an IVR-based training in order to arrive at a set of categorical evaluation/appraisal ratings. The fuzzy comprehensive evaluation model is established through the following steps (Chen *et al.*, 2015). The first step involves the establishment of the evaluation index system which is represented by a competency factor set $U = \{u_1, u_2, \cdots, u_n\}$ consisting of the factors that have a significant impact to the user performance in IVR-based training system where n is the number of the factors (Chen *et al.*, 2015).

The second step is the determination of the set of evaluation/appraisal grades (Chen *et al.*, 2015). These evaluation/appraisal grades are represented as a vector $V = \{v_1, v_2, \cdots, v_m\}$, in which *m* represents the number of stages in the appraisal/evaluation standards (Chen *et al.*, 2015). Generally, the evaluation/appraisal grade for each stage will be determined by an expert panel.

The third step involves the creation of the fuzzy evaluation matrix, R (Chen *et al.*, 2015). Typically, a fuzzy subset R_i in a competency factor set U is represented as $R_i = (r_{i1}, r_{i2}, \cdots, r_{im})$ where $r_{i1}, r_{i2}, \cdots, r_{im}$ correspond to the degree of membership of the *i*-th factor while *m* represents the number of stages in the appraisal/evaluation standards and $r_{i1} + r_{i2} + \cdots + r_{im} = 1$ (Chen *et al.*, 2015). After establishing the individual fuzzy subsets of all factors in U, then it is possible to establish the fuzzy evaluation matrix R (Equation 4.1) expressed as the interrelation between the factor set U and the evaluation/appraisal set V (Chen *et al.*, 2015). In order to determine the membership function of each factor and the fuzzy evaluation matrix R, several geometric mapping functions such as triangular or trapezoidal mapping functions can be used (Liu *et al.*, 2020).

$$R = \begin{bmatrix} R_1 \\ \vdots \\ R_n \end{bmatrix} = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nm} \end{bmatrix} \qquad \textit{Equation 4.1}$$

The fourth step involves the creation of the weight vector W consisting of weights of evaluation factors in a competency factor set U (Chen *et al.*, 2015). For instance, the weight for n competency factors can be represented by the vector $W = (w_1, w_2, \cdot\cdot, w_n)$ where $w_1, w_2, \cdot\cdot, w_n$ are weights of n competency factors, n is the number of factors and $w_1 + w_2 + \cdots + w_n = 1$ (Chen *et al.*, 2015). To obtain a comprehensive evaluation, the relative importance of each competency factor in the overall grading should be quantified. The weight vector can be formulated through pairwise comparison method, ranking method, point allocation method, or mean weight method (Ma *et al.*, 2010; Odu, 2019).

The last step involves the establishment of the overall appraisal result (Chen *et al.*, 2015). Through taking into consideration the weights of evaluation criteria, the overall appraisal result can be obtained. The evaluation vector B is calculated using the formula $B = W^\circ R = (b_1, b_2, \cdots, b_m)$ where, W is the weight vector, R is the fuzzy evaluation matrix, and $b_1, b_2, \cdots, b_m$ represent the comprehensive membership degree values to each evaluation/appraisal standard (Chen *et al.*, 2015).

4.5. Proposed assessment framework

This assessment framework involves the development of competency-based assessment rubrics in the IVR setting (Stage I), the validation of the developed assessment rubrics (Stage II), and the evaluation of overall user performance score (Stage III). Figure 4.1 shows the flow of the assessment framework in IVR setting.

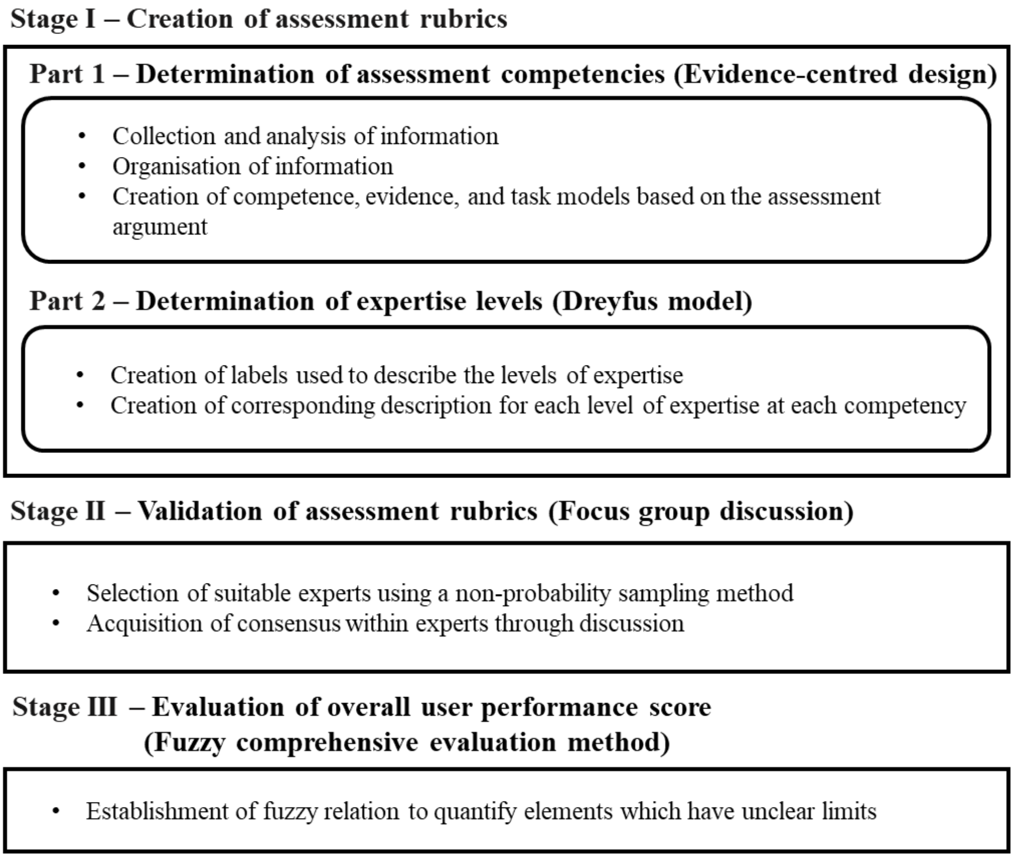


Figure 4.1. Proposed flow of the framework of user expertise evaluation in IVR setting.

In order to establish the competency-based assessment rubrics in IVR setting (Stage I), this study employed a combination of several methodologies such as the ECD model, and the Dreyfus model. The ECD methodology was integrated to identify the appropriate competencies to be rated and to organise evidentiary interrelations between the chain of reasoning and the collected information in IVR environment, while the Dreyfus model was adopted to create a baseline for identifying the current level of expertise of a trainee and to determine the appropriate instructions or feedback for the needs of the trainee (Dreyfus, 2004; Mislevy *et al.*, 2003).

To verify the authenticity of the developed IVR-based assessment rubrics (Stage II), this study used a focus group discussion with a group of experts to explore their views on a particular topic and to draw a group consensus (O.Nyumba *et al.*, 2018). After validation, the overall user performance score is evaluated using a fuzzy comprehensive evaluation method. The developed assessment framework was implemented in the IVR training called ‘Operate Your Own Reactor’ (OYOR) (Tehreem *et al.*, 2022).

4.6. Case Study – IVR training for production of *n*-butyllithium (*n*-BuLi)

4.6.1. Overview of the prototype

This study aims to integrate the proposed assessment framework in ‘Operate Your Own Reactor’ (OYOR), an IVR-based safety training system on the operational procedures and/or processes of the production of *n*-butyllithium (*n*-BuLi or *n*-C_4H_9Li) (Garcia Fracaro *et al.*, 2021). OYOR (Figure 4.2) was created in Unity3D software along with XR toolkit for enabling VR interactions of operators to a simulated chemical plant. It runs on the Oculus Quest virtual reality head mounted display with Oculus Touch controllers and it is available in English, French, and German (Tehreem *et al.*, 2022). OYOR was created by four PhD students from different universities including the author of this thesis. The other students used OYOR to study how the VR technology can improve learning outcomes, to explore the use of VR for the advanced training of employees, and to understand the effectiveness and efficiency of VR training when applied to actual industrial operations. The overall contents as well as the design of OYOR were developed in collaboration with industrial experts to ensure authenticity.

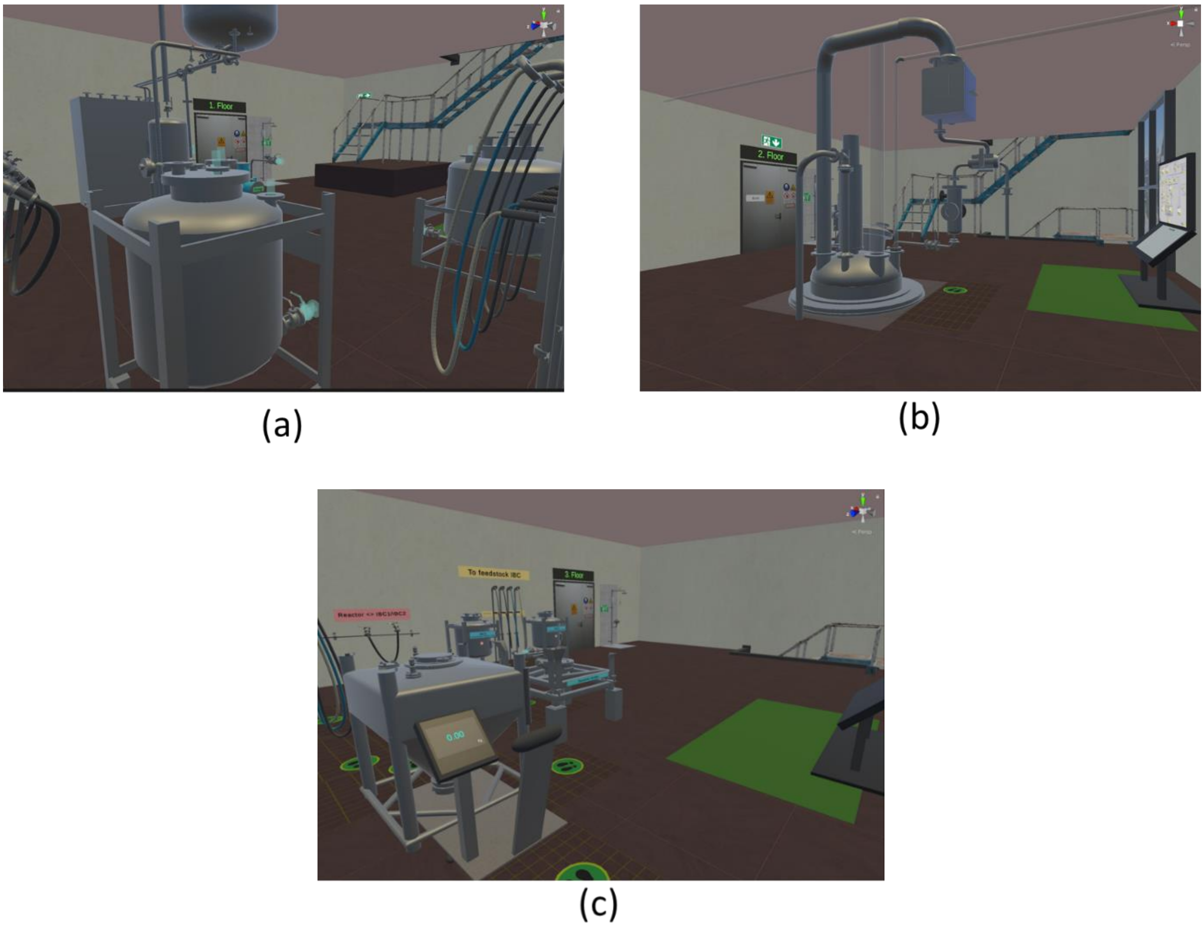


Figure 4.2. VR environment of the IVR-based training system used (a) 1st floor (b) 2nd floor and (c) 3rd floor.

The entire training takes approximately 60 minutes to complete, and consists of four steps: preparation, set-up, reaction, and extraction, which comprise 24 tasks of the *n*-BuLi production procedure (Figure 4.3) (Tehreem *et al.,* 2022). In order to verify whether the intended learning outcomes for every task and/or sub-task were achieved, operators need to demonstrate several skills such as how to operate and choose correct action from the computer screen, how to perform electrical grounding of an equipment, how and when to open or close certain valve, *etc*. in the IVR environment. In order to ensure that the users were comfortable and the side-effects of the usage of VR-HMD devices were minimised for more sensitive users, they were given an option to take a break after finishing each of the four steps.

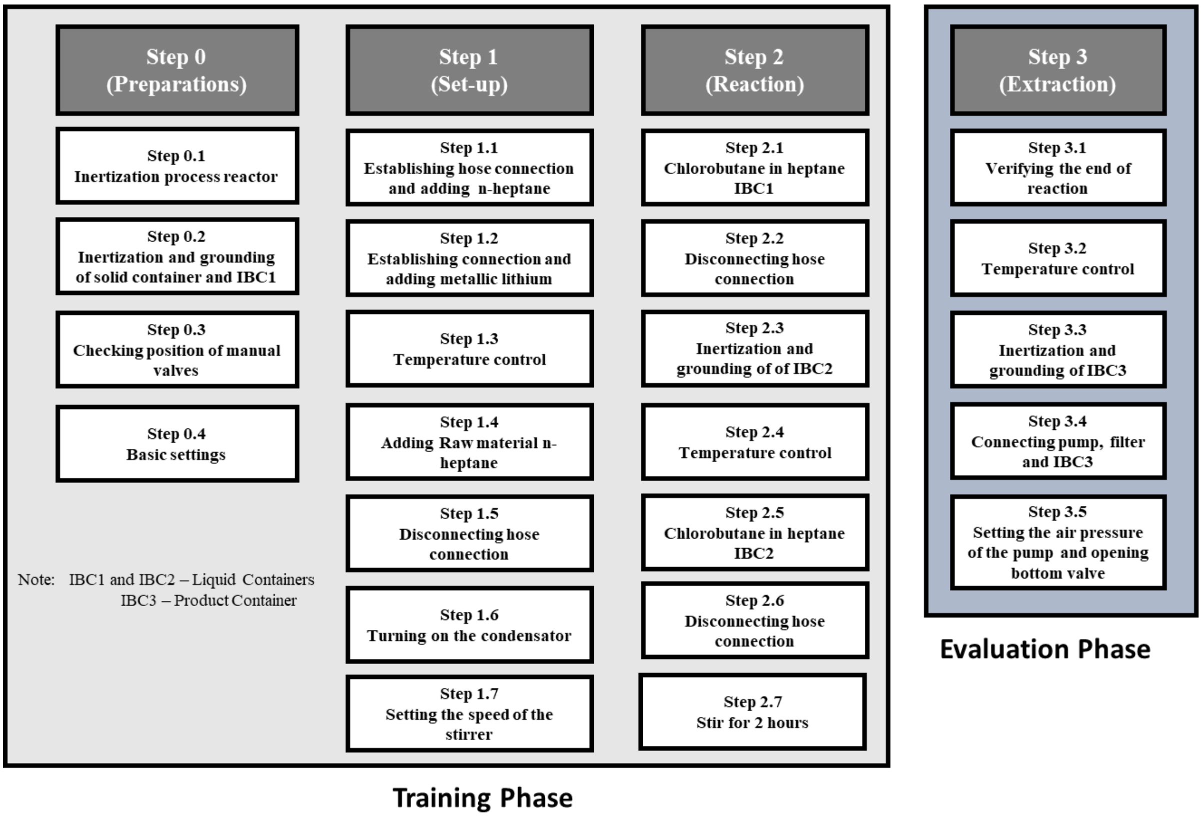


Figure 4.3. Procedural Layout of the IVR-based *n*-BuLi training system used.

Two types of interactions were implemented in this IVR-based training system: (1) actual interaction with the tools inside the VR environment, and (2) control of the reactor through a computer screen inside the VR environment (Tehreem *et al.,* 2022). As this IVR-based training was designed to ensure that the users can demonstrate and make use of their prior knowledge (*e.g.* how and when to connect/disconnect hosepipes, how to verify certain condition in IVR environment, *etc*.) they learned in subsequent *n*-BuLi production procedure, the first three steps, namely the preparation, set-up, and reaction were assigned as a ‘training phase’ (Figure 4.4). During this phase, users received detailed instructions and offered an automatic built-in support in a form of hints which directed learners to make correct actions. The last step, extraction, was assigned as an ‘evaluation phase’ (Figure 4.5), where the instructions were provided in a general way and hints were provided when requested by the users.

Figure 4.4. Screenshots of the instructions and hints in the training phase.

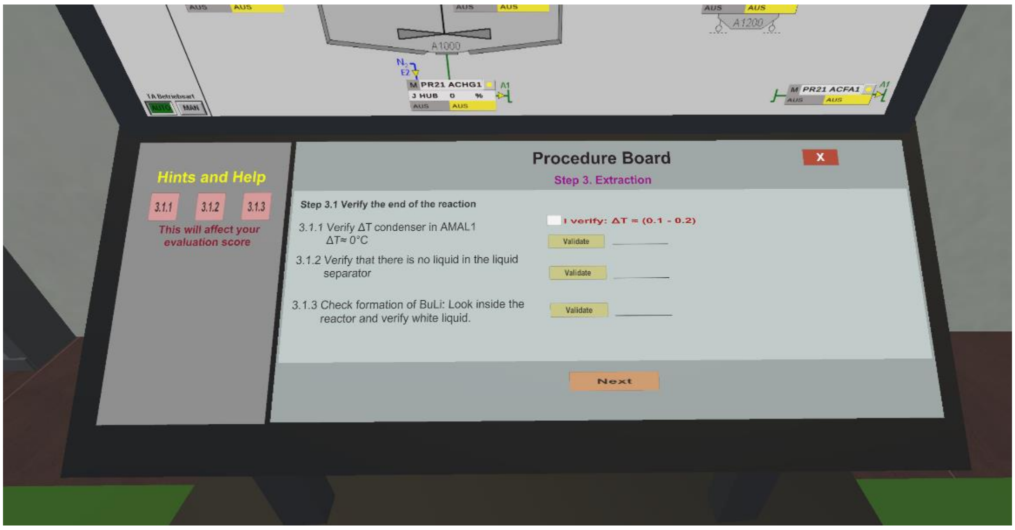


Figure 4.5. Screenshot of the instructions and hints in the evaluation phase.

4.6.2. Walkthrough the application of IVR-based assessment framework

As the goal of the prototype is to measure user performance and to gauge the competency level of the trainees, the actions of a trainee are recorded in a form of easily retrievable files during the session. In order to create an automated process of measuring competency level of the trainees for operational and safety procedures during *n*-BuLi production, the framework of user expertise evaluation described above, was employed as explained in the sub-sections below.

4.6.2.1. Creation of assessment rubrics

It is important to develop assessment rubrics to accurately measure and evaluate user performance criteria such as knowledge, skills, and/or attitude. Rubrics are assessment tools, often in a form of a matrix, that can help trainers to come to a similar conclusion about construction of certain performance criteria based on the scaled levels of mastery/expertise development that are indexed to an appropriate standard (Allen & Tanner, 2006). Typical rubrics are composed of three elements (Brookhart, 2013). The first element are the criteria used to enumerate and describe the characteristics to be rated (*e.g.* knowledge, skills, and/or behaviour). The second element are the standards used to categorise the levels of expertise development. The last element are the descriptors which provide a detailed characterisation of each criterion at each standard.

Since the success or failure of assessment rubrics in IVR setting depends on gathering and selecting the relevant information from the given log data, it is important to establish the means of collecting valid information and providing a substantial argument that connects what trainees do, say, or create in a given scenario. In order to formulate a list of qualities that the trainees should demonstrate so it is possible to evaluate their progress in the hierarchy of skills based on the pieces of evidence from task performances in IVR, ECD is adopted in this research.

The ECD consists of three models namely a competency model, an evidence model, and a task model. These models are designed to address a series of questions that are crucial in any assessment design. The competency model identifies and defines trainee-related variables (*e.g.* knowledge, skills, *etc.*), the inclusion of which are justified by conducting a literature review and/or series of brainstorming activities (Mislevy *et al.*, 2003). In building a competency model, it is important to specify the variables clearly in order to establish their interrelationships. As Tynjälä (2008) argued, theoretical knowledge (*i.e.* knowledge of particular facts), practical knowledge (*i.e.* procedural knowledge of how to perform task or simply as skills), and self-regulative knowledge (*i.e.* knowledge guided by metacognition, strategic action, and motivation to learn) are the three basic elements that are closely interrelated in the development of professional expertise. Through brainstorming and categorising every essential variable in the IVR environment, it is possible to create a blueprint about targeted aspects of capabilities. Table 4.2 shows the proposed competency model for the evaluation phase of the OYOR prototype.

Table 4.2. Proposed competency model for the evaluation phase of the OYOR prototype describing specific learning outcomes of the training.

Variables	Learning outcomes
Practical knowledge	To demonstrate how to connect/disconnect a hosepipe to a given equipment.
	To demonstrate how to perform grounding of a given equipment.
	To demonstrate how to open/close a ball valve.
	To verify the given condition in the IVR setting.
	To demonstrate how to choose a correct action in the main computer screen.
	To solve the given level in a minimum timeframe.
Self-regulative knowledge	To solve the given stage with minimum help/hints

The execution of a specific task significantly contributes to provide evidence of the performance of a trainee *vis-à-vis* his/her target competency. Hence, the task model in the assessment framework identifies what activities, including their specific features and conditions, should be performed to assess the competencies of trainees (Mislevy *et al.*, 2003). A task model is composed of various tasks designed to establish, among others, what the learner will be asked, the kind of response they are allowed, and the type of formats available. For instance, the verification of the end of the *n*-BuLi reaction process involves three sub-tasks namely verifying the temperature change in AMAL1 tank, verifying if the reflux ended, and verifying the formation of *n*-BuLi. Moreover, each sub-task contains specific actions that would help assess the behaviours and the performance of the trainees. The detailed list of tasks during the evaluation phase, as well as the proposed task model for the OYOR prototype, are shown in Table 4.3.

Table 4.3. Proposed task model for the evaluation phase (Main Task 4 – Extraction) of the OYOR prototype.

Main tasks	Sub-tasks	Specific actions	Instrument Involved
Complete all tasks of verifying the end of reaction. (3.1)	Complete the task of verifying the temperature change in AMAL1 tank.	• Search for the AMAL1 panel on the control monitor. • Observe the screen and check if the temperature change is 0 °C.	Control screens
	Complete the task of verifying if the reflux ended.	• Search for the liquid separator. • Observe the separator and check if the liquid stops flowing	Liquid separator

	Complete the task of verifying the formation of BuLi.	• Search for the main reactor. • Observe the inside of the reactor through the glass and check if white liquid is present.	Reactor
Complete the task of temperature control. (3.2)	Complete the task of performing action on the control screen.	• Search for the ALAC1 panel on the control monitor. • Click the drop box option under inner temperature and set to "20 °C" • Choose the start button and click it.	Control screens
Complete the task of Inertisation. (3.3)	Complete the task of IBC3 container connection.	• Search for the IBC3 container on the VR environment. • Search for the nitrogen pipe source and connect to the designated area. • Search for the exhaust air pipe source and connect to the designated area. • Search for the electrical grounding source and connect to the designated area. • Search for the level indicator and connect to the designated area.	IBC3 container
Complete the task of grounding and connecting pump/filter. (3.4)	Complete the task of connecting clamp for electric charges.	• Search for the filter in the VR environment. • Search for the electrical grounding source and connect to the designated area (Filter).	Filter
	Complete the task of connecting clamp for electric charges.	• Search for the pump in the VR environment. • Search for the electrical grounding source and connect to the designated area (Pump).	Pump
	Complete the task of connecting hose from the reactor and to the pump.	• Search for the hose • Connect the hose to the reactor and to the pump • Search for the location of the ball valve connecting pump and filter	Hose pipe and ball valve

		• Turn on the ball valve	
	Complete the task of connecting the hose from pump to the filter.	• Search for the pipe/hose source • Connect the hose from pump to the filter • Search for the location of the ball valve connecting pump and filter • Turn on the ball valve	Hose pipe and ball valve
	Complete the task of connecting the hose from filter to IBC3.	• Search for the hose • Connect the hose from filter to IBC3 • Search for the location of the ball valves connecting filter and IBC3 • Turn on all ball valves	Hose pipe and ball valve
Complete the task of air pressure setting for the pump. (3.5)	Complete the task of ACHG1 to docking station of IBC3.	• Search for the ACHG1 panel on the control monitor. • Click the drop box option under Receiver and set to “AAC71”. • Click the drop box option under Operation mode and set to “Empty”. • Choose the start button and click it. • Search for the AAC71 panel on the control monitor. • Click the drop box option under Operation mode and set to “befull reaktor”.	Control screens
	Complete the task of opening valve (digitally) using ACHG1.	• Search for the ACHG1 panel on the control monitor. • Click the drop box option under Sollwert (set point) and set to “100%”. • Choose the start button and click it.	Control screens
	Complete the task of activating air supply to pump from ACFA1.	• Search for the ACFA1 panel on the control monitor.	Control screens

		• Search for the valve "PR21 ACFA1 MFH01" and select open.

The evidence model acts as an intermediary to interrelate the competency and task models (Mislevy *et al.*, 2003). It shows how the outcomes of the actions that come from the interaction of the trainees with a given task are converted to observable variables, and it shows how these observable variables convey some information about the target criterion. Table 4.4 shows the relationship between several competences, evidence, and tasks for the evaluation phase of the OYOR prototype.

Table 4.4. Relationship between various competences, evidence, and tasks for the evaluation phase of the OYOR prototype.

Competence	Evidence	Task
Practical knowledge	• Number of mistakes committed	• Search for the "_____" hosepipe and connect/disconnect to the designated area. • Turn on/off the "____" valve. • Verify the liquid is flowing into the "____" equipment. • Search for the "____" on the control screen and perform the correct action.
Practical knowledge	• Total time of completion	• Complete Level ___ tasks.
Self-regulative knowledge	• Number of hints requested	• Complete all tasks in Level ___ with least number of hints.

After identifying the list of criteria through the ECD framework, the standards (or levels) of expertise development need to be separately categorised and described. As most training designers face the challenge of developing a curriculum that explains how trainees move up from one level to another, this study adopted the Dreyfus model to categorise the levelling up process of trainees. Through creating suitable expertise development scheme for employees training during *n*-BuLi production, it will be easier for the trainers to determine the current level of expertise of their trainees and to develop the appropriate guidance needed for them to move to the next level.

Once the identification of the important criteria as well as the level of expertise development are completed, it is possible to develop the assessment rubrics with a detailed description of each criterion at each level of expertise development. Although benchmarking the specified level of expertise development is a challenging process, this task may help to guide trainees toward an insight that the assessment rubrics epitomise the principles of a line of work.

The categorisation of the Dreyfus model, as applied in this study has some variations from the original model (refer to Table 4.5). In the original Dreyfus model, as discussed in Chapter 4.2, the trainee moves up from a novice to expert level, with the expert defined as someone who is better in decision-making and has better sense of responsibility. This would normally apply in large areas of discipline relevant skills/knowledge, where the decision-making and sense of responsibility need to be practiced at higher level. Arguably, the scope of this study in its application in OYOR IVR-based training programme limits the 'responsibility' of the operator to taking the right sequence of actions at the right time. Thus, while this study uses the same labels as the original Dreyfus model ("novice", "advanced beginner", "competent", "proficient", and "expert"), they do not, strictly speaking, correspond to the same characteristics. However, the use of an easy-to-understand feedback system (*e.g.* same labels as the Dreyfus model) is beneficial to the trainees in order to motivate them to perform better prospectively, in reference to their past performance (Gallagher *et al.*, 2022). It is expected that once a trainee successfully completed the OYOR programme as an expert (or "competent" in the original Dreyfus model), such person may be qualified to be deployed as an operator in the specific chemical plant. It has to be noted, however, that under certain circumstances, *e.g.* particularly safety critical operations, it is essential that a trainee performs the tasks faultlessly. This can be easily accommodated by redefining the classification boundaries (*e.g.* a trainee would be classified as an expert only if the procedure was carried out without a single mistake and in a required timeframe). Whilst this was not the case in the OYOR case study, which served as a demonstrator in this thesis, the proposed framework is easily adapted to the requirements of a particular application.

Table 4.5. Original Dreyfus model *vis-à-vis* the Dreyfus model as applied in the OYOR VR training programme.

Original Dreyfus model (Dreyfus, 2004; Dreyfus & Dreyfus, 1987)	**Dreyfus model as applied in OYOR**
Novice	**Novice**
• Novices can only perform their tasks by following a set of guidelines. Thus, they have little situational perception and behave with restricted judgment. When	• **Novice** requires close supervision and assistance to perform their tasks. To perform without supervision, they require a specific set of guidelines.

confronted with complicated tasks, they need close supervision.	They have little situation perception and behave with restricted judgment.
Advanced Beginner • At this level, the advanced beginner has some appreciation of specific details about learning but still has limited situational perception. They are developing an ability to evaluate relevance of information but may still tend to rely on guidebooks. As such, they provide limited answers to unusual or complicated tasks since they do not always realise the possible results. They can perform regular sequence of tasks under indirect supervision, but still need it for complicated tasks.	**Advanced Beginner** • Advanced Beginner in OYOR environment is equivalent to an **"apprentice"** in industry standards. An advanced beginner requires supervision from time to time to ensure that their tasks are completed according to standards. For complicated tasks or unusual situation, they still require close supervision.
Competent • Competent individuals start to see how actions may affect long-term results. They can organise and assess the circumstances to concentrate on important details. They perform deliberate planning, understand procedures by experts, and can adapt to new settings. They can perform complex analytical tasks and planning but still need supervision for non-routine complicated tasks.	**Competent** • Individuals at this level are considered **"advanced beginner"** in the Dreyfus model. These individuals have appreciation of specific details about learning but still have limited situation perception. They are developing an ability to evaluate relevance of information but may still tend to rely on guidebooks. As such, they provide limited answers to unusual or complicated tasks since they do not always realise the possible results. They can perform regular sequence of tasks under indirect supervision, but still need it for complicated tasks.
Proficient • Individuals at this level are more accountable and confident and can deal with complicated scenarios holistically. They comprehend with clarity good and relevant information to resolve problems. At this stage, they are beginning to develop intuitive judgment and solve problems based on prior knowledge. Proficient individuals understand rules, theories, and alternatives more deeply.	**Proficient** • Proficient individuals in OYOR environment are normally considered **"qualified"** in industry standards. Neitzel (2018), stated that a qualified person is an individual had training and knowledge to perform a specific task. Citing Occupational Safety and Health Administration (OSHA), he noted that an individual may be considered qualified in one specific task, such as operating a particular equipment, but not necessarily to another related task (*e.g.* other equipment) (Neitzel, 2018). While

	they have an understanding of relevance of their tasks, they do not yet have the foresight to understand their long-term effects.
Expert • Experts are better in decision-making through intuitive reasoning. They use analytical approaches in adapting to new situations, and they no longer rely on policies and procedures. They have a sense of responsibility for themselves, others, and the environment while envisioning the overall picture and possible alternatives.	**Expert** • Experts in this study are considered **"competent"** in the Dreyfus model. Competent individuals start to see how actions may affect long-term results. They can organise and assess the circumstances to concentrate on important details. They perform deliberate planning, understand procedures by experts, and can adapt to new settings. They can perform complex analytical tasks and planning but still need supervision for non-routine complicated tasks.

Assessment rubrics give trainers a structured means to decide at what level of expertise a trainee is currently and to determine what type of task, instruction, and/or feedback they need. In the OYOR setting, for instance, an individual must complete the task without relying on the hints in order to move from proficient to expert. Table 4.6 shows the proposed assessment rubrics for the progression of an individual from novice to expert for the evaluation phase of the OYOR prototype.

Table 4.6. Proposed assessment rubrics for determining the current level of expertise for the evaluation phase of the OYOR prototype.

Criteria	**Description**				
	Novice	Advanced Beginner	Competent	Proficient	Expert
Number of mistakes (action) committed	Complete the task with more than 10 mistakes.	Complete the task with 6-10 mistakes.	Complete the task with 3-5 mistakes.	Complete the task with 1-2 mistakes.	Complete the task without mistakes.
Total time of completion	Take more than 10 minutes compared to expected time to complete the task, or does not complete	Complete the task within 6-10 minutes of the expected time.	Complete the task within 3-5 minutes of the expected time.	Complete the task within 1-2 minutes of the expected time.	Complete the task ahead of the expected time per scenario.

Number of hints requested	Complete the task with the help of more than 10 hints.	Complete the task with the help of 6-10 hints.	Complete the task with the help of 3-5 hints.	Complete the task with the help of 1-2 hints.	Complete the task without using hints.

4.6.2.2. Validation of assessment rubrics

Since the formulated assessment rubrics for IVR environment must be comparable with what is being used in the real world, it is important to verify the developed IVR-based assessment rubrics in order to gauge their ability to provide sufficient and clear interpretations, conclusions, and recommendations related to the learning and development of trainees. For this study, the validation of assessment rubrics was conducted through an online group discussion using Microsoft Teams with four experts, health and safety trainers and supervisors, working at MERCK KGaA. This is an international science and technology company located in Darmstadt, Germany and an industry beneficiary of the Horizon 2020 funded European Training Network for Chemistry Engineering Immersive Learning project (Grant Agreement 812716). Before conducting the validation process, ethics approvals were obtained from the Ethics Committee at the university as well as at the industrial partner institution. All expected ethical procedures were followed during the online group discussion.

The discussion process began with presenting the prototype as well as the proposed assessment rubrics. The experts then provided feedback on the assessment rubrics. For instance, the experts recommended that for the evaluation phase, instead of collecting and presenting the total number of mistakes, it would be more beneficial if the system could present this information in such a way that both the users and the trainers know where the mistakes occurred, *i.e.,* collecting mistakes based on the action they did on computer screen, opening/closing valves, and connecting/disconnecting hose. After considering the expert recommendations, the updated version of the assessment rubrics for determining the current level of expertise from step 3.1 to step 3.5 of the evaluation phase are shown in Table 4.7, Table 4.8, Table 4.9, Table 4.10, and Table 4.11, respectively.

Table 4.7. Assessment rubrics for determining the current level of expertise for the step 3.1 of the evaluation phase.

Main criteria	**Sub-criteria**	**Description**				
		Novice	Advanced Beginner	Competent	Proficient	Expert

Practical knowledge	Number of mistakes (Visual verification)	Complete the task with more than 10 mistakes.	Complete the task with 6-10 mistakes.	Complete the task with 3-5 mistakes.	Complete the task with 1-2 mistakes.	Complete the task without mistakes.
	Total time of completion	Take more than 11.51 minutes to complete the task, or does not complete	Complete the task within 6.51 – 11.5 minutes.	Complete the task within 3.51 – 6.5 minutes.	Complete the task within 1.51 – 3.5 minutes.	Complete the task in less than or equal to 1.5 minutes
Self-regulative knowledge	Number of hints requested	Complete the task with the help of more than 3 hints.	Complete the task with the help of 3 hints.	Complete the task with the help of 2 hints.	Complete the task with the help of 1 hint.	Complete the task without using hints.

Table 4.8. Assessment rubrics for determining the current level of expertise for the step 3.2 of the evaluation phase.

Main criteria	**Sub-criteria**	**Description**				
		Novice	Advanced Beginner	Competent	Proficient	Expert
Practical knowledge	Number of mistakes (Control screen work)	Complete the task with more than 10 mistakes.	Complete the task with 6-10 mistakes.	Complete the task with 3-5 mistakes.	Complete the task with 1-2 mistakes.	Complete the task without mistakes.
	Total time of completion	Take more than 10.51 minutes to complete the task, or does not complete	Complete the task within 5.51 – 10.5 minutes.	Complete the task within 2.51 – 5.5 minutes.	Complete the task within 0.51 – 2.5 minutes.	Complete the task in less than or equal to 0.5 minutes
Self-regulative knowledge	Number of hints requested	Complete the task with the help of more than 3 hints.	Complete the task with the help of 3 hints.	Complete the task with the help of 2 hints.	Complete the task with the help of 1 hint.	Complete the task without using hints.

Table 4.9. Assessment rubrics for determining the current level of expertise for the step 3.3 of the evaluation phase.

Main criteria	**Sub-criteria**	**Description**				
		Novice	Advanced Beginner	Competent	Proficient	Expert

Practical knowledge	Number of mistakes (Inertisation)	Complete the task with more than 10 mistakes.	Complete the task with 6-10 mistakes.	Complete the task with 3-5 mistakes.	Complete the task with 1-2 mistakes.	Complete the task without mistakes.
	Total time of completion	Take more than 11.01 minutes to complete the task, or does not complete	Complete the task within 6.01 – 11 minutes.	Complete the task within 3.01 – 6 minutes.	Complete the task within 1.01 – 3 minutes.	Complete the task in less than or equal to 1 minutes
Self-regulative knowledge	Number of hints requested	Complete the task with the help of more than 3 hints.	Complete the task with the help of 3 hints.	Complete the task with the help of 2 hints.	Complete the task with the help of 1 hint.	Complete the task without using hints.

Table 4.10. Assessment rubrics for determining the current level of expertise for the step 3.4 of the evaluation phase.

Main criteria	**Sub-criteria**	**Description**				
		Novice	Advanced Beginner	Competent	Proficient	Expert
Practical knowledge	Number of mistakes (Hoses connection)	Complete the task with more than 10 mistakes.	Complete the task with 6-10 mistakes.	Complete the task with 3-5 mistakes.	Complete the task with 1-2 mistakes.	Complete the task without mistakes.
	Number of mistakes (Electrical grounding)	Complete the task with more than 10 mistakes.	Complete the task with 6-10 mistakes.	Complete the task with 3-5 mistakes.	Complete the task with 1-2 mistakes.	Complete the task without mistakes.
	Number of mistakes (Operation of valves)	Complete the task with more than 10 mistakes.	Complete the task with 6-10 mistakes.	Complete the task with 3-5 mistakes.	Complete the task with 1-2 mistakes.	Complete the task without mistakes.
	Total time of completion	Take more than 14.01 minutes to complete the task, or does not complete	Complete the task within 9.01 – 14 minutes.	Complete the task within 6.01 – 9 minutes.	Complete the task within 4.01 – 6 minutes.	Complete the task in less than or equal to 4 minutes

Self-regulative knowledge	Number of hints requested	Complete the task with the help of more than 3 hints.	Complete the task with the help of 3 hints.	Complete the task with the help of 2 hints.	Complete the task with the help of 1 hint.	Complete the task without using hints.

Table 4.11. Assessment rubrics for determining the current level of expertise for the step 3.5 of the evaluation phase.

Main criteria	**Sub-criteria**	**Description**				
		Novice	Advanced Beginner	Competent	Proficient	Expert
Practical knowledge	Number of mistakes (Control screen work)	Complete the task with more than 10 mistakes.	Complete the task with 6-10 mistakes.	Complete the task with 3-5 mistakes.	Complete the task with 1-2 mistakes.	Complete the task without mistakes.
	Total time of completion	Take more than 11.51 minutes to complete the task, or does not complete	Complete the task within 6.51 – 11.5 minutes.	Complete the task within 3.51 – 6.5 minutes.	Complete the task within 1.51 – 3.5 minutes.	Complete the task in less than or equal to 1.5 minutes
Self-regulative knowledge	Number of hints requested	Complete the task with the help of more than 3 hints.	Complete the task with the help of 3 hints.	Complete the task with the help of 2 hints.	Complete the task with the help of 1 hint.	Complete the task without using hints.

4.6.2.3. Calculation of the overall user performance score

After identifying and validating the most important information in the IVR environment, it is important to retrieve this information in a form of easily retrievable files to evaluate the overall performance of the trainees. In order to achieve this, fuzzy comprehensive evaluation method was adopted. As discussed in section 4.4, this method is composed of 5 steps. The first step is the establishment of the evaluation index system. To structure an overall user performance evaluation index in the OYOR prototype, performance measures such as practical knowledge and self-regulative knowledge were selected as competency factor set *U* such that appropriate measurements (*e.g.* number of hints, total time, number of mistakes, *etc.*) were collected to evaluate the performance level of the learners based on these factors. Figure 4.6 shows the developed evaluation index for user overall performance in the OYOR prototype.

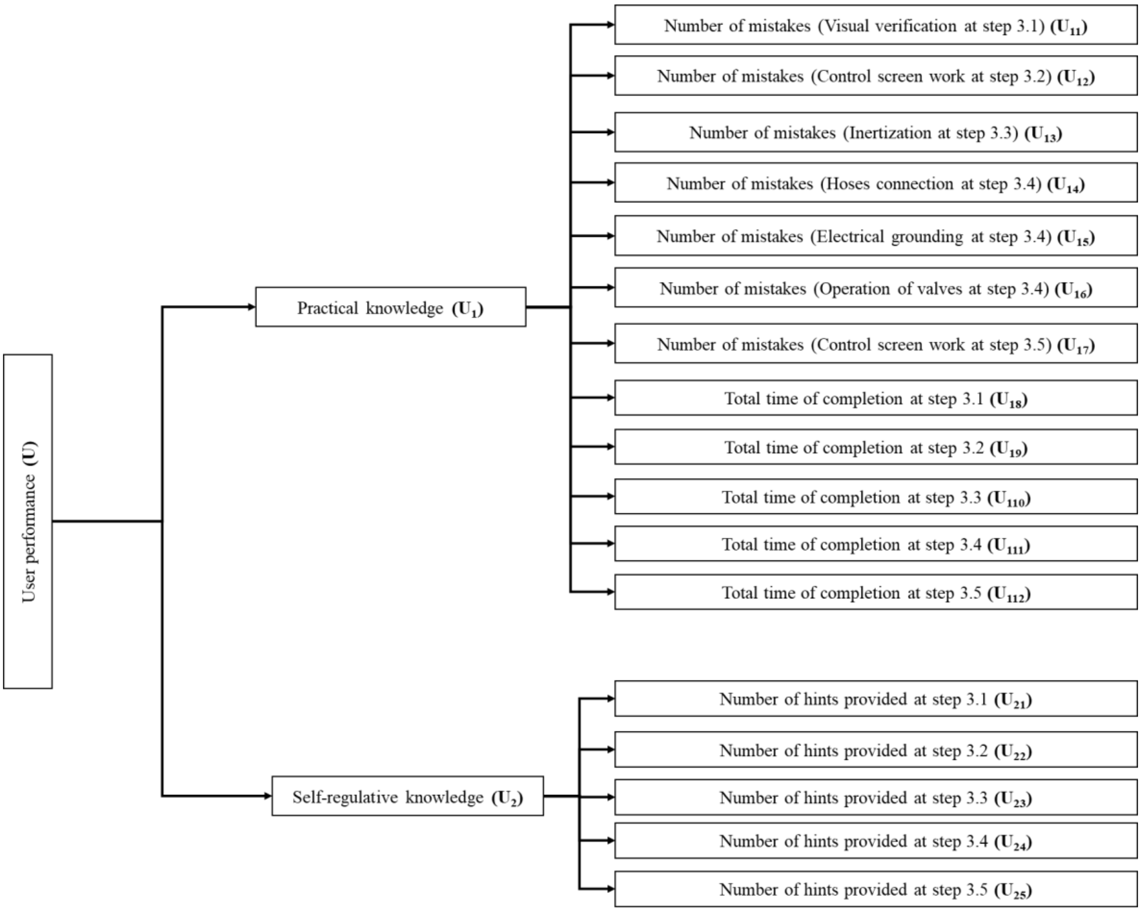


Figure 4.6. Evaluation index for user overall performance in the evaluation phase of OYOR prototype.

In order to determine the mapping quality for the overall user performance in the IVR environment, it is important to determine a set of appraisal grades (step 2). For this study, the same experts working at MERCK KGaA were asked to comment on the proposed appraisal grade values for every performance level in the developed assessment rubrics. Table 4.12 shows the proposed appraisal scale of the overall user performance level in the OYOR prototype.

Table 4.12. Proposed scale of the user overall performance level.

Performance Level	**I**	**II**	**III**	**IV**	**V**
Note	Novice	Advanced beginner	Competent	Proficient	Expert
Region	[0%-40%)	[40%-60%)	[60%-80%)	[80%-90%)	[90%-100%]
Appraisal Score	0%	40%	60%	80%	100%

After discussion, the experts recommended to change the appraisal grade values for the overall user performance scale. The new appraisal grade scale were shown in Table 4.13. The standard appraisal grade values for every measured action in the IVR environment is shown in Table 4.14.

Table 4.13. Validated scale of the user overall performance level.

Performance Level	**I**	**II**	**III**	**IV**	**V**
Note	Novice	Advanced beginner	Competent	Proficient	Expert
Region	[0%-50%)	[50%-67%)	[67%-81%)	[81%-92%)	[92%-100%]
Appraisal Score	0%	50%	67%	81%	100%

Table 4.14. Standard appraisal grade values for every measured action in the OYOR prototype.

Competency factors (units)	**Performance Level**				
	I	**II**	**III**	**IV**	**V**
U_{11} (point)	11	10	5	2	0
U_{12} (point)	11	10	5	2	0
U_{13} (point)	11	10	5	2	0
U_{14} (point)	11	10	5	2	0
U_{15} (point)	11	10	5	2	0
U_{16} (point)	11	10	5	2	0
U_{17} (point)	11	10	5	2	0
U_{18} (min)	11.51	11.5	6.5	3.5	1.5
U_{19} (min)	10.51	10.5	5.5	2.5	0.5
U_{110} (min)	11.01	11	6	3	1
U_{111} (min)	14.01	14	9	6	4
U_{112} (min)	11.51	11.5	6.5	3.5	1.5
U_{21} (point)	4	3	2	1	0
U_{22} (point)	4	3	2	1	0
U_{23} (point)	4	3	2	1	0
U_{24} (point)	4	3	2	1	0
U_{25} (point)	4	3	2	1	0

The next step involves the establishment of the fuzzy membership function of the evaluation matrix, *R*. For this study, a triangle-shape grade of the membership function is adopted as this is the most common fuzzy membership function used by several authors such as (M. kun Li *et al.*, 2018) and (Lai *et al.*, 2015). Equations 4.2 to 4.6 show the general fuzzy membership function $r_{ij}(x_i)$ related to the five expertise/performance levels. The symbol r_{ij} represents the membership values of the *j*-th performance level of the *i*-th factor in U while the symbol x_i and $P_{i,j}$ represents the actual value of the *i*-th factor and the maximum of the *j*-th performance level (j = 1–5) of the *i*-th factor, respectively. Through using these formulas, it is possible to change ambiguity to certainty and to solve the problems of ambiguity and the nonlinear relation between the membership value and level accurately.

$$r_{i,1}(x_i) = \begin{cases} 1 & x_i \geq P_{i5} \\ \dfrac{x_i - P_{i4}}{P_{i5} - P_{i4}} & P_{i4} < x_i < P_{i5} \\ 0 & x_i \leq P_{i4} \end{cases} \qquad \textit{Equation 4.2}$$

$$r_{i,2}(x_i) = \begin{cases} 0 & x_i \leq P_{i3} \text{ } or \text{ } x_i \geq P_{i5} \\ \dfrac{x_i - P_{i3}}{P_{i4} - P_{i3}} & P_{i3} < x_i < P_{i4} \\ \dfrac{P_{i5} - x_i}{P_{i5} - P_{i4}} & P_{i4} \leq x_i \leq P_{i5} \end{cases} \qquad \textit{Equation 4.3}$$

$$r_{i,3}(x_i) = \begin{cases} 0 & x_i \leq P_{i2} \text{ } or \text{ } x_i \geq P_{i4} \\ \dfrac{x_i - P_{i2}}{P_{i3} - P_{i2}} & P_{i2} < x_i < P_{i3} \\ \dfrac{P_{i4} - x_i}{P_{i4} - P_{i3}} & P_{i3} \leq x_i \leq P_{i4} \end{cases} \qquad \textit{Equation 4.4}$$

$$r_{i,4}(x_i) = \begin{cases} 0 & x_i \leq P_{i1} \text{ } or \text{ } x_i \geq P_{i3} \\ \frac{x_i - P_{i1}}{P_{i2} - P_{i1}} & P_{i1} < x_i < P_{i2} \\ \frac{P_{i3} - x_i}{P_{i3} - P_{i2}} & P_{i2} \leq x_i \leq P_{i3} \end{cases} \qquad \textit{Equation 4.5}$$

$$r_{i,5}(x_i) = \begin{cases} 0 & x_i \geq P_{i2} \\ \dfrac{P_{i2} - x_i}{P_{i2} - P_{i1}} & P_{i1} < x_i < P_{i2} \\ 1 & x_i \leq P_{i1} \end{cases} \qquad \textit{Equation 4.6}$$

For instance, if one finishes a task involving Inertisation (***U13***) with 4 mistakes (x_i), then applying equations 4.2 to 4.6, the evaluation matrix, ***R13*** = [r_{131} r_{132} r_{133} r_{134} r_{135}] = [0 0 0.67 0.33 0]. The sample calculation is shown below:

$$r_{131}(4) = \begin{cases} 1 & x_i \geq 11 \\ \dfrac{x_i - 10}{11 - 10} & 10 < x_i < 11 \\ 0 & x_i \leq 10 \end{cases}$$

$$r_{131}(4) = 0$$

$$r_{132}(4) = \begin{cases} 0 & x_i \leq 5 \; or \; x_i \geq 11 \\ \dfrac{x_i - 5}{10 - 5} & 5 < x_i < 10 \\ \dfrac{11 - x_i}{11 - 10} & 10 \leq x_i \leq 11 \end{cases}$$

$$r_{132}(4) = 0$$

$$r_{133}(4) = \begin{cases} 0 & x_i \leq 2 \; or \; x_i \geq 10 \\ \dfrac{x_i - 2}{5 - 2} & 2 < x_i < 5 \\ \dfrac{10 - x_i}{10 - 5} & 5 \leq x_i \leq 10 \end{cases}$$

$$r_{133}(4) = \frac{4 - 2}{5 - 2} = 0.67$$

$$r_{134}(4) = \begin{cases} 0 & x_i \leq 0 \; or \; x_i \geq 5 \\ \dfrac{x_i - 0}{2 - 0} & 0 < x_i < 2 \\ \dfrac{5 - x_i}{5 - 2} & 2 \leq x_i \leq 5 \end{cases}$$

$$r_{134}(4) = \frac{5 - 4}{5 - 2} = 0.33$$

$$r_{135}(4) = \begin{cases} 0 & x_i \geq 2 \\ \dfrac{2 - x_i}{2 - 0} & 0 < x_i < 2 \\ 1 & x_i \leq 0 \end{cases}$$

$$r_{135}(4) = 0$$

Through this, every membership degree of competency factor *i* to appraisal grade *j* can be calculated and thus the evaluation matrix for the overall user performance (***R***) as well as for the competency such as practical knowledge ($\boldsymbol{R_1}$) and self-regulative knowledge ($\boldsymbol{R_2}$) can be obtained. Table 4.15 shows the actual data and the membership values of competency factors for a certain user.

Table 4.15. Actual values and membership values of competency factors.

Competency factors (unit)	**Actual values**	**Membership values**				
		I	**II**	**III**	**IV**	**V**
U_{11} (point)	0	0	0	0	0	1
U_{12} (point)	0	0	0	0	0	1

U_{13} (point)	4	0	0	0.67	0.33	0
U_{14} (point)	1	0	0	0	0.5	0.5
U_{15} (point)	5	0	0	1	0	0
U_{16} (point)	0	0	0	0	0	1
U_{17} (point)	1	0	0	0	0.5	0.5
U_{18} (min)	1.38	0	0	0	0	1
U_{19} (min)	0.3	0	0	0	0	1
U_{110} (min)	2.22	0	0	0	0.61	0.39
U_{111} (min)	8.2	0	0	0.73	0.27	0
U_{112} (min)	4.07	0	0	0.19	0.81	0
U_{21} (point)	0	0	0	0	0	1
U_{22} (point)	0	0	0	0	0	1
U_{23} (point)	2	0	0	1	0	0
U_{24} (point)	2	0	0	1	0	0
U_{25} (point)	0	0	0	0	0	1

After establishing the evaluation matrix, it is vital to determine the weighing vector as this gives an overview of the proportion of each competency factor in the overall user performance index based on relative importance (step 4). According to Odu (2019), mean weight method is the most widely adopted weighing method if there is no information available from the experts. Given that the OYOR prototype is new and there is insufficient information available to decide which of the abovementioned actions is more important, the mean weight (equal importance) method is adopted for this study. To determine the mean weight, equation 4.7 is used. The list of the competency factor weights for determining the overall user performance in the OYOR prototype is shown in Table 4.16.

$$w_1 = w_2 = \cdots = w_{17} = \frac{1}{number\ of\ competency\ factors} \quad \textit{Equation 4.7}$$

$$= \frac{1}{17}$$

$$= 0.059$$

Table 4.16. Overall user performance index weights for OYOR prototype.

Competency	Weight	Competency factors	Relative weight	Absolute weight
U_1	0.706	**U_{11}**	0.083	0.059
		U_{12}	0.083	0.059
		U_{13}	0.083	0.059
		U_{14}	0.083	0.059
		U_{15}	0.083	0.059
		U_{16}	0.083	0.059
		U_{17}	0.083	0.059
		U_{18}	0.083	0.059

		U_{19}	0.083	0.059
		U_{110}	0.083	0.059
		U_{111}	0.083	0.059
		U_{112}	0.083	0.059
		U_{21}	0.200	0.059
		U_{22}	0.200	0.059
U_2	0.294	**U_{23}**	0.200	0.059
		U_{24}	0.200	0.059
		U_{25}	0.200	0.059

To determine the overall user performance result (step 5), the formula $B = W^\circ R$ is used. Using the values of Table 4.15, the evaluation matrix for the overall user performance (***R***) of a certain user is summarised as:

$$R = \begin{bmatrix} R_{11} \\ R_{12} \\ R_{13} \\ \vdots \\ R_{25} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0.67 & 0.33 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

To calculate the evaluation vector $B_{Overall}$;

$$B_{Overall} = W^\circ R$$

$$B_{Overall} = [0.059 \quad 0.059 \quad 0.059 \quad \ldots \quad 0.059]^\circ \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0.67 & 0.33 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B_{Overall} = [0 \quad 0 \quad 0.270 \quad 0.178 \quad 0.552]$$

In order to convert the evaluation vector to a single score which represent their overall performance level during the training in IVR environment, the obtained evaluation vector ($B_{Overall}$) is multiplied by the appraisal score set found at Table 4.13 $[0 \quad 50 \quad 67 \quad 81 \quad 100]^T$. Applying this;

$$User\ performance_{Overall} = [0 \quad 0 \quad 0.270 \quad 0.178 \quad 0.552][0 \quad 50 \quad 67 \quad 81 \quad 100]^T$$

$$User\ performance_{Overall} = 88\% = Proficient$$

4.6.2.4. Implementation of assessment framework

In order to understand how the validated assessment rubrics can be used by different stakeholders to utilise the information for their respective accountability purposes, the validated rubrics were then implemented in the evaluation phase of the OYOR prototype. After the implementation, testing of the prototype was carried out with operator apprentices from ACTA, a chemical training company located in Brasschaat, Belgium, over a 2-day period (November 2021). This company was also an industrial partner of the Horizon 2020 funded European Training Network for Chemistry Engineering Immersive Learning project (Grant Agreement 812716). Before the data collection, an ethical authorisation was obtained from the Ethics Committee at the respective institutions.

Apprentices were invited to voluntarily take part in the testing of the IVR-based training prototype. Prior to the start of the experiment, participants who agreed to take part were asked to complete the pre-training questionnaire which consisted of a written informed consent and closed questions used to determine demographic data such as gender and age. The demographic information of the participants is summarised in Table 4.16. The majority of the participants were male (89.47%) and between the ages of 20-30 (84.21%).

Table 4.17. Demographic information of participants (n=19).

Characteristics	**Items**	**Frequency**	**Percentage**
Gender	Male	17	89.47
	Female	1	5.26
	No answer/Prefer not to say	1	5.26
Age	18-19	2	10.53
	20-30	16	84.21
	No answer/Prefer not to say	1	5.26

After answering the pre-training questionnaire, the participants were given a brief introduction as well as a demonstration on the use of the Oculus Quest 2 device. Upon completion of the introductory session, the participants were instructed to begin the *n*-BuLi production process in IVR environment. At the end of the evaluation phase of the IVR training, the summary of the assessment system (Figure 4.7) was shown for every participant.

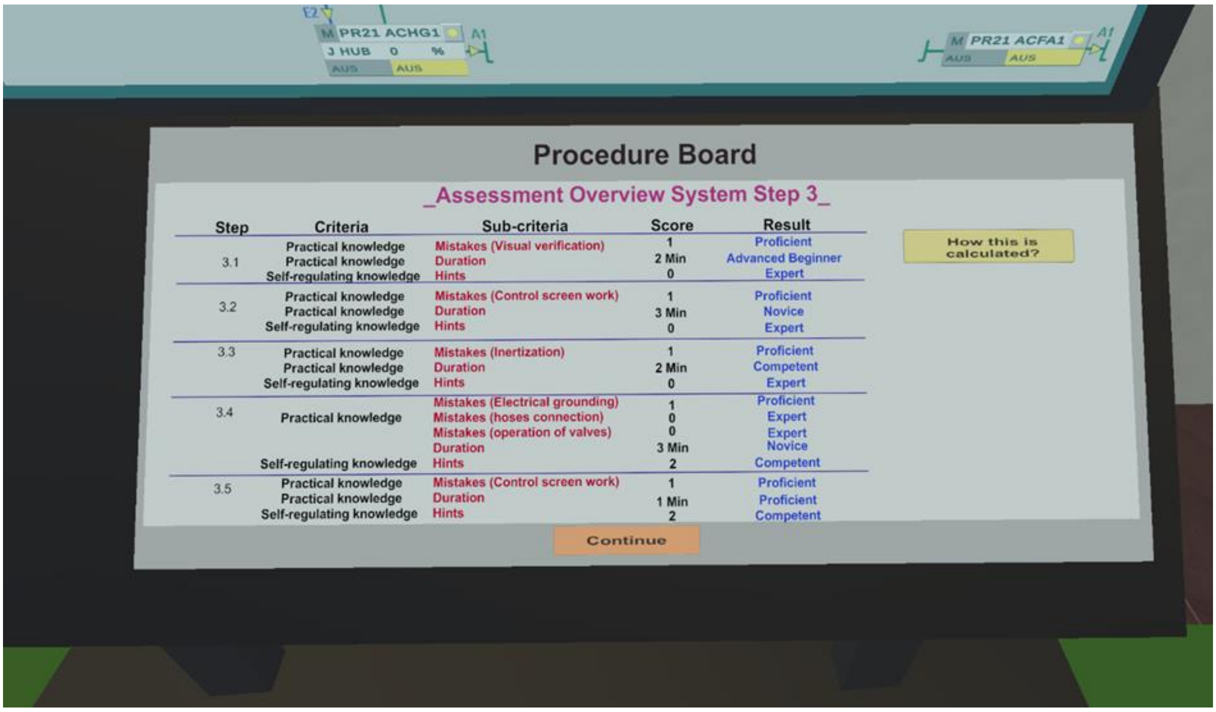


Figure 4.7. Screenshot of the assessment overview system in the evaluation phase.

Given that the summary of user performance data (Figure 4.7) was recorded in a form of easily retrievable .csv files, it is possible to compile all the data to create a visualisation reporting tool which helps to track, analyse, and display data about the overall status of participants in terms of their skills at every task in the evaluation phase of the IVR training. Figures 4.8, 4.9, 4.10, 4.11 and 4.12 show the visualisation report for categorising the performance level of 19 participants in the task 3.1, 3.2, 3.3, 3.4, and 3.5 of the evaluation phase, respectively. The overall performance remark of the 19 participants is shown in Figure 4.13.

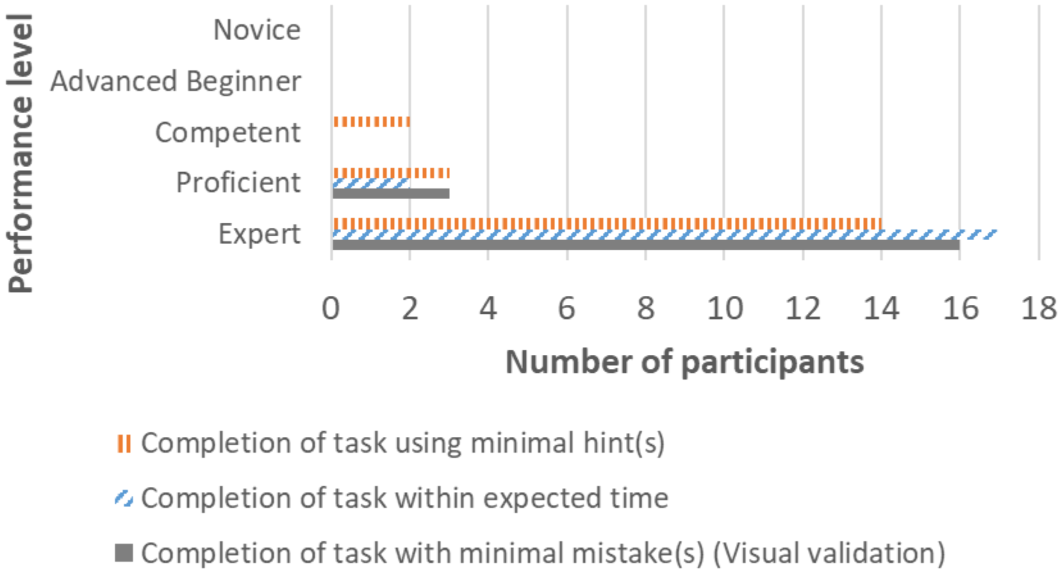


Figure 4.8. Example of visualisation report for categorising the performance level of 19 participants in the task 3.1 of the evaluation phase.

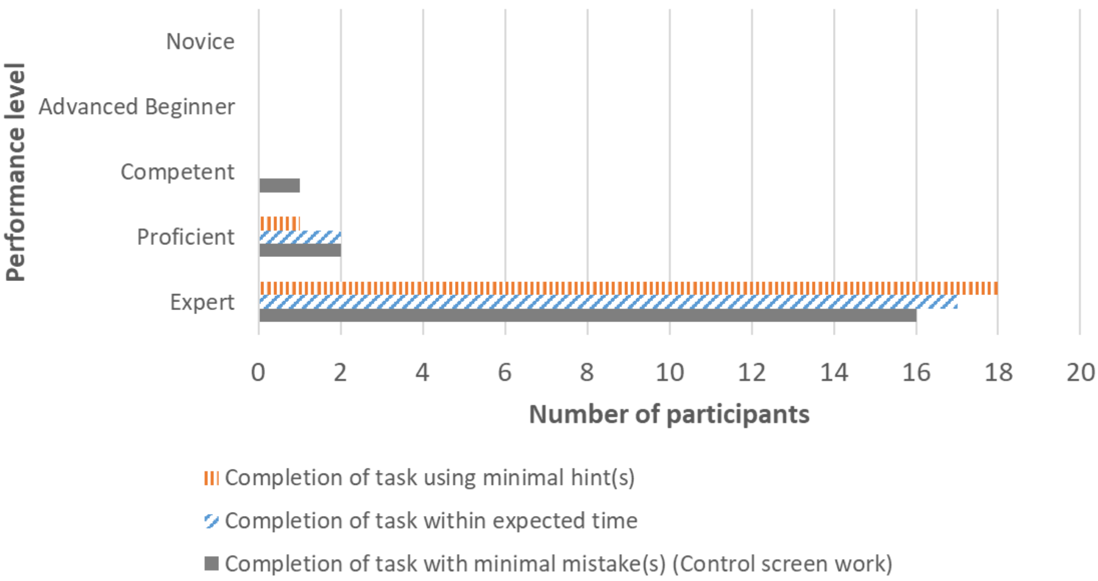


Figure 4.9. Example of visualisation report for categorising the performance level of 19 participants in the task 3.2 of the evaluation phase.

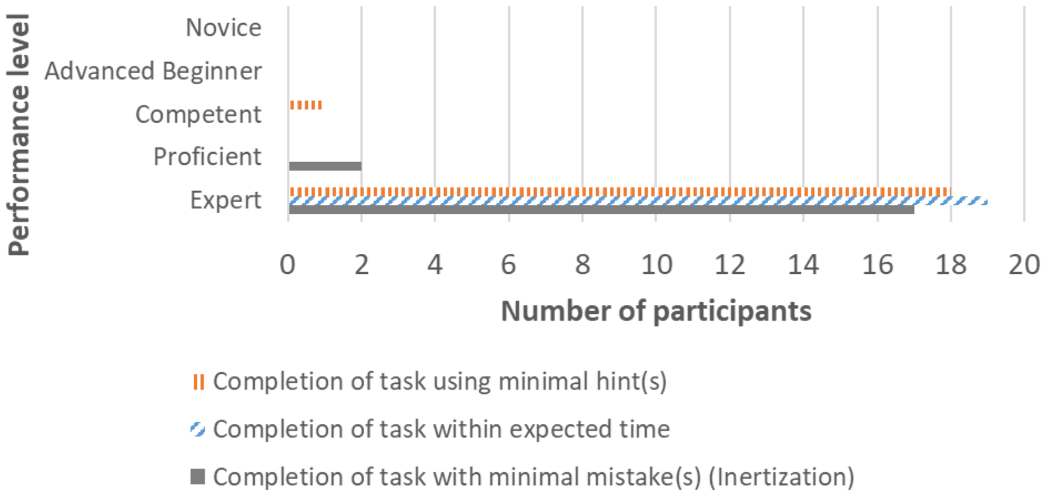


Figure 4.10. Example of visualisation report for categorising the performance level of 19 participants in the task 3.3 of the evaluation phase.

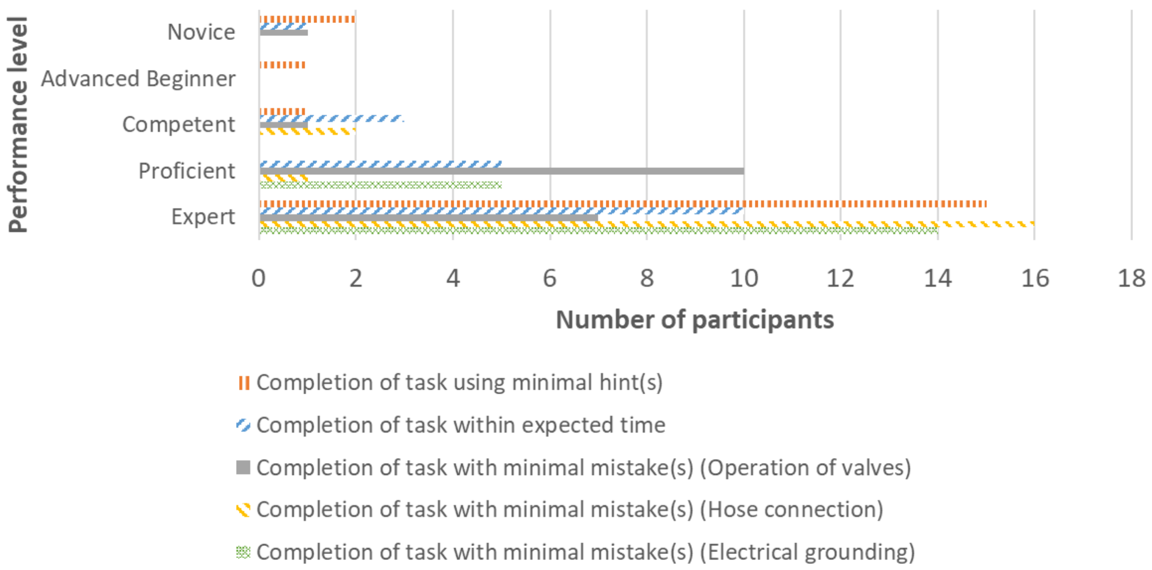


Figure 4.11. Example of visualisation report for categorising the performance level of 19 participants in the task 3.4 of the evaluation phase.

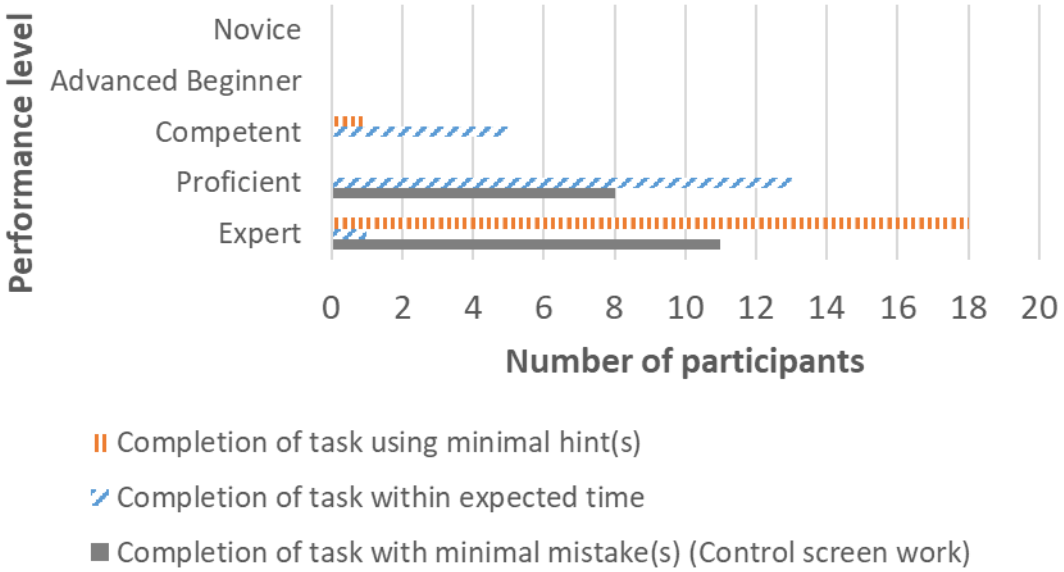


Figure 4.12. Example of visualisation report for categorising the performance level of 19 participants in the task 3.5 of the evaluation phase.

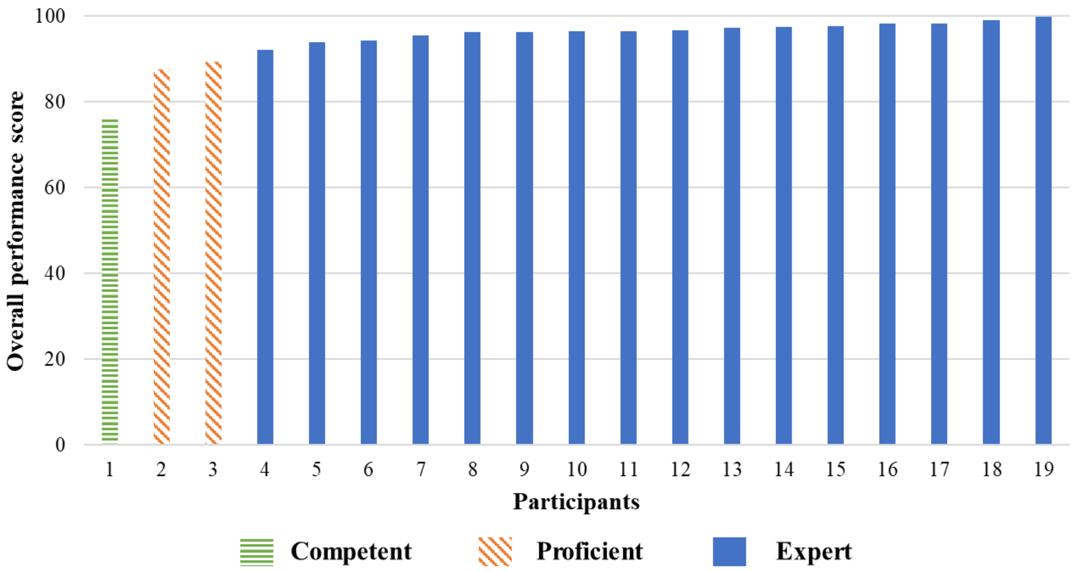


Figure 4.13. Overall performance rating of 19 participants in the evaluation phase.

4.7. Implications

The integration of ECD, Fuzzy comprehensive evaluation method, and Dreyfus model into an assessment framework powered through IVR has broad implications for various stakeholders, especially in high-risk industries. Primarily, this study will facilitate the conduct of training involving complex and possibly intricate procedures, in a safe and reliable manner as chemical operator apprentices need to be equipped and trained in its potentially hazardous environment. The transparency in the conduct of training and assessment methods also increases the integrity of the training outcomes, which in turn serves as a trustworthy reference for all stakeholders in their individual decision-making.

Trainees, in particular, will be able to learn by experience, allowing them to advance their competency towards mastery and expertise of a particular function or task. In addition, they will be able to track and monitor their training progress through a more detailed and accurate feedback in the IVR training. This will enable them to pinpoint their strengths and weaknesses in the accomplishment of their responsibilities. The information and data provided to the trainee during the IVR training may then be used by them to readily review and correct their mistakes and misconceptions, if any. This continuous process of improvement will eventually lead to a more effective knowledge and skill acquisition. Lastly, trainees will be able to objectively evaluate their status in terms of skills, within the organisation, which they may use as leverage and basis for promotion.

The application of this study will also greatly benefit the trainers. This will minimise the use of discretion in the evaluation process resulting in more objective judgments as the assessment rubrics are already embedded in the training system. This will also help the trainers in independently evaluating employees for the purposes of personnel movement with the objective of ensuring an efficient working environment. Moreover, this assessment framework will also assist the trainers in providing a more constructive and personalised feedback for trainees to ensure their continual growth.

Since the resulting framework of this study is designed to expedite the development of experts in a specific field, administrators may appreciate a shorter and more predictable return-on-investment (ROI) in terms of personnel development. The presence of more personnel who are proficient, competent, and experts in the field will result a more productive and efficient working environment that will significantly benefit the business. Thus, the integration of this assessment framework in IVR environment will allow the administrators to determine the overall state and condition of the organisation in terms of field expertise.

4.8. Conclusions

The development of experts in a specific field requires a progressive process of skill and knowledge acquisition. This theory is set forth in the Dreyfus model describing the growth from a novice through an advanced beginner, to a competent, then a proficient, and finally, to an expert level. The model provides characteristics and measures to differentiate each stage in the expert development.

In facilitating the training of experts through a safe, reliable, and accurate method, the Dreyfus model may be optimised by integrating this concept into an IVR training environment. A training design in IVR allows the users an in-depth experience of the actual work scenario with lower risk than the traditional training methods.

In this set-up, a large amount of information and data will be available to the trainers. To determine the relevant information and data out of all available inputs as well as to produce accurate inferences, the ECD is introduced in this framework. ECD facilitates the systematic collection of data and relates these to theoretical constructs through evidentiary reasoning.

After identifying and validating the most important information in the IVR environment, the fuzzy comprehensive evaluation method is implemented to evaluate the overall performance of the users. As this method allows to quantify factors which have no clear borders (*e.g.* linguistic scores), this will be helpful for the stakeholders to know the evaluation 'score'.

The entire framework of the integration of Dreyfus model and fuzzy comprehensive evaluation method in the IVR setting with assessment carried out through ECD was tested and implemented in a case study. The case study involves an IVR-training named 'Operate Your Own Reactor' where 19 chemical operator apprentices from ACTA, a chemical training company based in Belgium, were trained to perform tasks for the production of *n*-butyllithium in a virtual chemical plant. The actions of the trainees were recorded during the session. Through the ECD blueprint, data such as the number of mistakes committed, total time of completion, and number of hints requested were retrieved and consolidated to assess the users. The data were analysed to identify the performance levels of the trainees including their current levels in the Dreyfus model. These collected performance results were subsequently used to determine the overall performance score of the user through fuzzy comprehensive evaluation method.

In summary, this framework provides detailed assessment guidelines for the unbiased and unobtrusive evaluation of expertise in immersive learning environments. Through this method, trainees and trainers will be able to determine, with specificity, their strengths and weaknesses in the performance of their tasks. In addition, administrators will be provided with relevant and timely information about the performance of plant personnel for their strategic decision-making in terms of manpower management and development. Ultimately, stakeholders will benefit in the long-term by expediting the development of experts leading to more cost-effective and efficient industry operations.

4.9. Note

Part of this chapter has been submitted (September 2022) in the Assessment in Education: Principles, Policy & Practice Journal: Toyoda, R., Garcia-Fracaro, S., Gallagher, T., Tehreem, Y., Russo Abegão, F., & Glassey, J. A framework for evaluating user expertise in immersive learning environments.

Chapter 5. Incorporating assessment into chemical industry training through immersive virtual reality (IVR): a case study

Chapter 4 detailed the integration of Dreyfus model in the immersive virtual reality (IVR) setting with assessment carried out through the evidence-centred design (ECD) and the fuzzy comprehensive evaluation methods in an IVR-based chemical process safety training called 'Operate your own reactor' (OYOR). The proposed assessment framework provides an unbiased and unobtrusive way of training and assessing chemical operator apprentices. To further develop the use of the OYOR prototype in chemical operators training, the presentation of feedback and assessment results in a manner that is acceptable and helpful to its potential users must be examined. This chapter aims to determine the overall preference of a user as to the details presented in the assessment results. The methodology employs both quantitative and qualitative methods to analyse the responses of 63 operator apprentices in the OYOR prototype. Apprentices were grouped into: (1) the control group, which received a simplified version of their performance, and (2) the experimental group, which received a detailed version of their performance. The quantitative analysis showed that there is no significant difference in attitude, satisfaction based on content, and behavioural intention in IVR training environment. The qualitative analysis further showed that the ability of the assessment system to ultimately provide performance evaluation is the primary concern for the apprentices. Both groups demonstrated positive attitude, satisfaction based on content, and behavioural intention. Nonetheless, in terms of user preference, most users in both groups prefer detailed assessment results to be provided in order to further review and improve their performance. The implications of IVR assessment were also discussed.

5.1. Introduction

The use of digital-based educational and/or training resources provides a wide range of potential for improving the efficacy and efficiency of knowledge, and skills transfer and development for different industries (Toyoda *et al.*, 2021). For instance, Domínguez Alfaro *et al.* (2022) used Augmented Reality (AR) to teach high school and undergraduate students about the concept of acid-based titration. Other examples include the use of a mobile game application to review organic reactions (da Silva Júnior *et al.*, 2021), the use of a web-based 360° virtual laboratory to teach undergraduate students regarding chemical laboratory safety (Viitaharju *et al.*, 2021), and the use of a virtual reality (VR) head-mounted display (HMD) to train professional chemical operators about industry-based chemical laboratory safety (Chan *et al.*, 2021). With this increased level of technology-assisted learning, the state of the art of assessment has been extensively investigated over the past decades to guarantee that the learners have the necessary 21st century skills such as problem solving, and communication skills, after undergoing a digital-based training (Black & Wiliam, 1998; Van der Kleij *et al.*, 2015).

According to Chin *et al.* (2009, p554), *"assessment describes the process of using data to demonstrate that stated goals and objectives are actually being met"*. Depending on how it is implemented, assessment can be classified into either summative (assessment of learning) or formative (assessment for learning) (Loh, 2012). As stated by Sadler (1989, p120), *"Summative assessment is concerned with summing up or summarising the achievement status of a student, and geared towards reporting at the end of a course of study especially for purposes of certification"* while *"formative assessment is concerned with how judgements about the quality of student responses (performances, pieces, or works) can be used to shape and improve the students' competence"*.

Researchers, such as Black & Wiliam (1998) and Baleni (2015), claim that the formative assessment plays a vital role in improving the effectiveness of learning since it involves users in understanding their current learning state and planning what they need to do next in order to develop their skills. Whilst the benefits of formative assessment are widely recognised, there are some challenges that need to be overcome. Some authors, for example Bennett (2011), pointed out that the process of creating and providing timely and high-quality judgements and/or comments with specific, detailed and personalised feedback tailored to the needs of increasing learner-educator ratios and more diverse learner profiles, results in a substantial increase in terms of the time and effort spent by the trainers and lecturers. Here the technology may offer some solutions. A number of institutions globally have been adopting the concept of digital assessment. According to Gomez and Ruipérez-Valiente (2022, p200), *"digital assessment is the presentation of evidence, for judging student achievement, obtained through the use of computer technology"*. Digital assessment can be categorised either automated assessment, semi-automated assessment, or non-automated/manual assessment. The automated assessment automatically evaluates the performance of the learners based on the predefined assessment parameters and the corresponding marking scheme, which are typically integrated in the digital-based training (Souza *et al.*, 2016). Examples of this type are the automated assessment in short essay type assignments developed by Dumal *et al.* (2017), and the automated assessment in Massive Open Online Courses (MOOCs) developed by Ivaniushin *et al.* (2016). The semi-automated assessment, such as the semi-automated assessment of student programming assignments developed by Jackson (2000) and the Java programming assessment tool developed by Yusof *et al.* (2012), assess programming solutions of the students automatically but ask the teacher to manually check the programming codes and give corresponding feedback if the said code is different from the expected output (Souza *et al.*, 2016). The non-automated or manual assessment, such as the programming assessment tool developed by Dawson-Howe (1995), compiles and executes the submitted programming assignment on a computer device but manually assesses the solution of every student and manually provides corresponding marks and feedbacks (Souza *et al.*, 2016). Among these three types of digital assessments, the automated digital assessment gained a significant level of attention and became broadly recognised as a possible solution to the intensive demand on assessment resources highlighted above (Barra *et al.*, 2020).

There are several studies that have investigated the potential impact of an automated digital assessment on learning. For instance, Baleni (2015) evaluated the effect of the automated digital formative assessment using a virtual learning environment (VLE) called Blackboard. According to their study, the integration of automated and immediate formative feedback in an online objective tests, such as multiple-choice or true or false tests, creates an opportunity for the students to assess and re-assess their understanding of the respective subject matter, rather than just presenting an overall grade (Baleni, 2015). Moreover, Kerr *et al.* (2016) used an automated digital formative assessment system called eQuip (eQuestions for Understanding Integrated Physiology) to evaluate the performance of the students. Their results show that students who used eQuip outperformed those students who did not use it in their year-end examination as the said assessment system provides immediate and timely feedback which helps to become aware and improve their performances, reflect on their own mistakes, and set goals for their learning (Kerr *et al.*, 2016). Furthermore, Barana *et al.* (2019) stated that the use of automated formative assessment for learning mathematics online can help students to master problem solving procedures through interactive feedback while working at their own pace.

On the other hand, Yılmaz *et al.* (2020) examined the disadvantages of automated digital formative assessment in an online multiple-choice examination. Their results reveal that the motivation of some participants were low as they feel demoralised whenever they receive low scores from the automated formative assessment system. In addition, Hricko and Howell (2006) commented that an automated formative digital assessment might imposes an additional demand on users as it requires reading and utilisation of numerous screens of assessment feedback such as texts and figures which might be more challenging and tiring to some individuals.

Such heterogeneous findings suggest that the impact of assessment in various digital-based learning is not straightforward. Thus, further research is needed to understand the views of the users on how digital assessment should be implemented in various learning environments. This action-based research reports the results of a case study aiming to determine whether the chemical operator apprentices (users) preferred a simple assessment (*i.e.* a simple overall grade statement), or a more detailed assessment (*i.e.* additional information, such as the number of mistakes, or hints per sub-task, aside from the overall grade) of their current performance evaluation in a chemical production training based on immersive virtual reality (IVR). The results of this study will contribute to our understanding of how digital assessment can be effectively implemented in IVR.

5.2. Theoretical underpinning and hypothesis development

The experience of a user trained using a specific training method may be measured in terms of its effectiveness, feasibility, and/or usability. However, the overall appraisal of the user in these aspects may differ from their appraisal of each of its specific components. For instance, the main driver of satisfaction of users may be the degree of immersion, design, and form of the training method but they may find other components in need of improvement, such as the user experience, or the presentation of assessment results and feedback (Checa & Bustillo, 2019). Thus, to ensure the success of the IVR-based training, the perspective of the learner in terms of their appreciation of each component of the new training/assessment method is crucial.

In particular, assessment results in the IVR-based training may be analysed using three theoretical constructs: attitude, satisfaction based on content, and behavioural intention. The first construct, attitude, is the predisposition of a person towards a particular phenomenon or object, and it is composed of affective (*i.e.* unconscious feelings) and cognitive (*i.e.* conjectures and ideology) components (H.-H. Teo *et al.*, 2003). Several studies used the concept of attitude to measure the degree of liking that the participants have on various digital-based products such as the use of e-learning (Hussein, 2017), and radio-frequency identification-based library management system (Zainab *et al.*, 2018). The results of these studies showed that the users express positive attitude towards a given product or service if it leads to the attainment of their important values such as personal improvement. Given that the user attitude to the assessment system in IVR environment depends on the comprehensiveness of the assessment provided, it is expected that the difference in the presentation of the assessment results used in this study would lead to a difference in the degree of user attitude. Thus, it is hypothesised:

H1: *There is a significant difference in attitude between groups based on the presentation scheme of the assessment system.*

The second construct is the satisfaction based on content. Heilman and Brusa (2006) and Xiao and Dasgupta (2002), examined the effect of this construct on a computer-based system and an internet/web portal system, respectively. Their results provide evidence that the likelihood of an individual being contented with a particular product or service was directly proportional to the amount, as well as the completeness, of the information provided for a specific purpose (Heilman & Brusa, 2006; Xiao & Dasgupta, 2002). From this, the following hypothesis is proposed:

H2: *There is a significant difference in satisfaction (based on content) between groups based on the presentation scheme of the assessment system.*

The last construct is behavioural intention. In the context of information technology (IT), this construct is defined as the likelihood of the user to use and to recommend the new technology to others (Venkatesh *et al.*, 2012). Since understanding the concept of user intention is a significant predictor of their actual behaviour, Toyoda *et al.* (2021) and Merhi *et al.* (2019) used the behavioural intention construct to explain user acceptance of IVR-based health and safety (H&S) training in chemical industry setting and mobile banking, respectively. As confirmed by the abovementioned studies, it is expected that the likelihood of the learner using the new training method in the future will increase if such method is shown to significantly help the learner in achieving his/her objectives. Thus, it is hypothesised:

H3: *There is a significant difference in behavioural intention between groups based on the presentation scheme of the assessment system.*

5.3. Methodology

5.3.1. Research design

For this study, an IVR-based chemical process safety training, running on the Oculus Quest 2 virtual reality head-mounted display called 'Operate Your Own Reactor' (OYOR), was used to train users in the operational procedures of the production of n-butyllithium (n-BuLi or *n*-C_4H_9Li) (Garcia Fracaro *et al.*, 2021). The entire training takes approximately 60 minutes and consists of four steps: preparation, set-up, reaction, and extraction, which comprise 24 tasks.

The training was designed to ensure that users will be able to demonstrate and use their existing knowledge and apply it to the *n*-Buli production procedure. The overview of the OYOR prototype was discussed in the previous chapter (Chapter 4).

Given the limited and inconclusive research on how assessment should be presented to users in IVR setting, a *quasi*-experimental study was set up. The participants were divided into a control group and an experimental group. The control group performed the VR-based training and were shown a simple assessment system (Figure 5.1). The simple assessment system shows only the overall performance score and whether they passed or failed. On the other hand, the experimental group performed the same VR-based training with a detailed assessment system (Figure 5.2). This assessment system provides a more comprehensive and detailed results including number of mistakes, time spent to complete the task and number of requested hints for the main task and sub-tasks as well as their overall performance. These parameters were selected and validated based on their relevance on performance assessment through deliberation with the MERCK KGaA health and safety officers. Moreover, these parameters are gathered automatically in the IVR platform simultaneous to the training of the chemical operator apprentices.

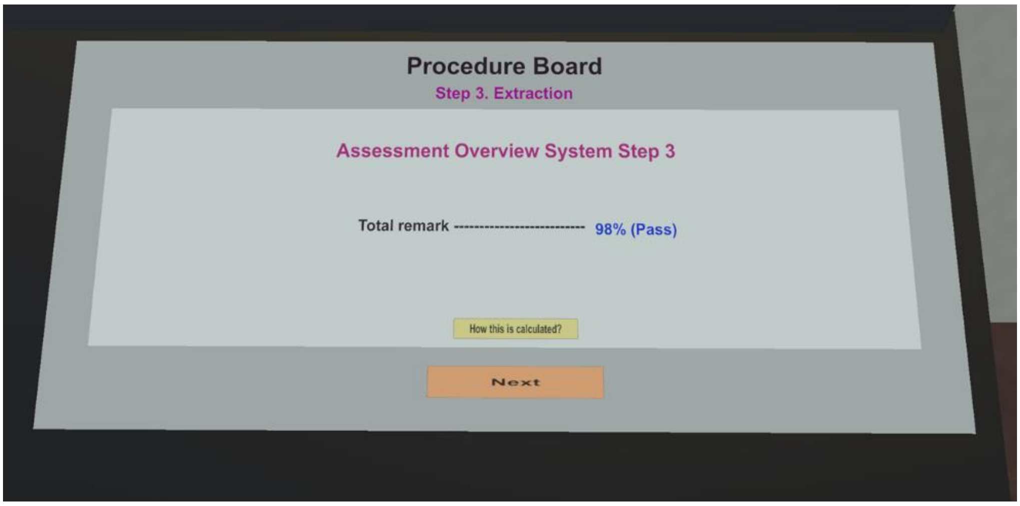


Figure 5.1. Screenshot of the simple assessment system.

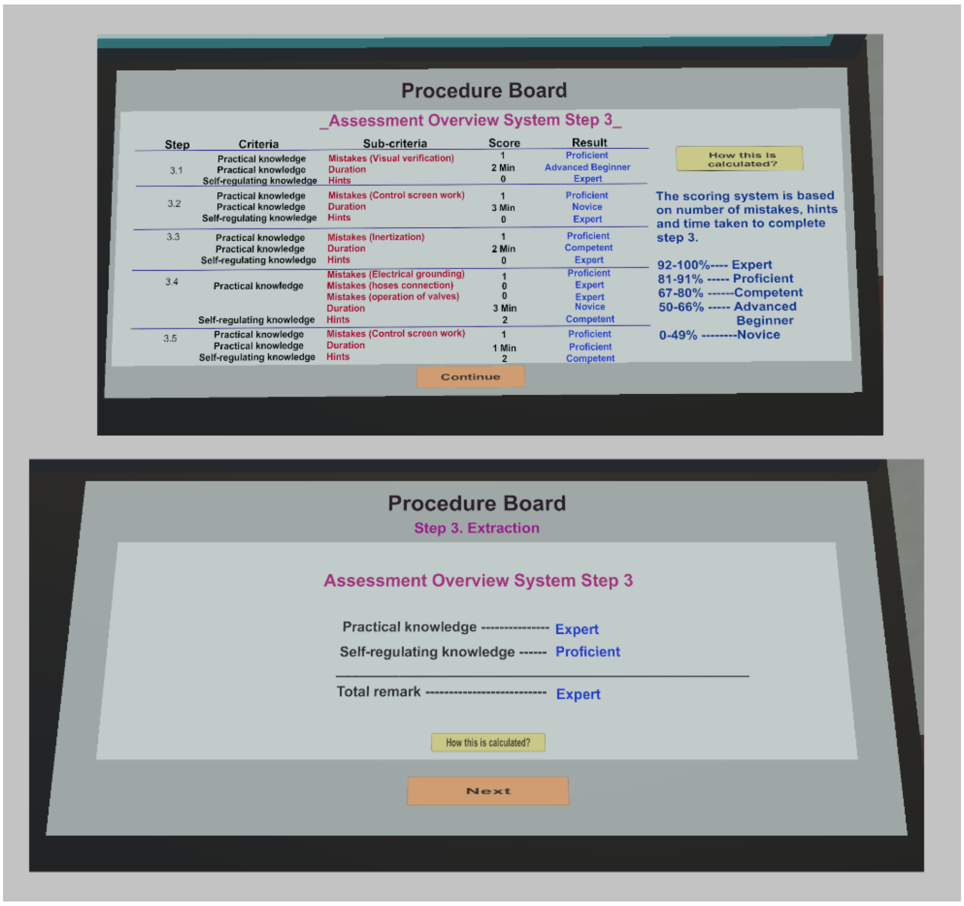


Figure 5.2. Screenshot of the detailed assessment system.

5.3.2. Questionnaire design

A three-part questionnaire was employed for this study. The first part was the pre-training questionnaire, which consisted of written informed consent and closed questions used to determine demographic data such as gender and age.

The second part was the post-training questionnaire, which comprised of two sections. The first section contained the items about attitude, satisfaction based on content, and behavioural intention that were adapted from previously reported literature and that were verified as valid and reliable (Marakarkandy & Yajnik, 2013; Venkatesh *et al.*, 2003). For each item, the questions were rephrased according to the scope of using assessment system in IVR training. The respondents indicated their agreement with each item about attitude and behavioural intention on a six-point Likert scale, ranging from 1 for strongly disagree to 6 for strongly agree. Each item about satisfaction based on content was evaluated on a number scale, ranging from 1 for never to 6 for always. The second section contained open-ended questions which asked respondents to identify advantages and/or disadvantages of the assessment system in IVR environment as these questions can elicit deeper insights on the use of assessment system in IVR environment that were not identified in the previous section.

The third part was the follow-up questionnaire, which consisted of two sections. The first section contained a closed question where a number of respondents from both groups were presented with two photos (Figure 5.1 and Figure 5.2) and were asked whether they would prefer the simple assessment system or the detailed assessment system. The second section contained an open-ended question where they were asked in what way it has affected their choice.

Face validity as well as content validity checks of the questionnaire were carried out by an academic expert (Dr Liesbeth Kester of Utrecht University, Netherlands). The modified questionnaire was then presented to a small number of voluntary professionals engineers working as health and safety (H&S) officers and/or trainers to evaluate its readability and comprehensiveness. The questionnaire was then revised according to their collective feedback. The final questionnaire was originally created in English, and subsequently translated into Dutch and German by a professional translator. To ensure the uniformity and validity of the translation, a separate professional translator then performed a blind back-translation of the questionnaire into English, which was compared with the initial English version (Dorer, 2012). The English list of items used in the study and corresponding constructs is shown in Table 5.1.

Table 5.1. Lists of measurement items used in the study.

Latent Variable	Item	Explanation
Attitude	ATT_1	Using the assessment overview system is a bad idea.*
	ATT_2	Using the assessment overview system is foolish.*
	ATT_3	Using the assessment overview system is favourable.
	ATT_4	Using the assessment overview system is beneficial.
Satisfaction based on content	Sat_1	Did the assessment overview system provide the precise information you need?
	Sat_2	Did the assessment overview system provide reports that seem to be just about exactly what you need?
	Sat_3	Did the assessment overview system provide sufficient information?
Behavioural Intention	BI_1	Assuming I had access to the assessment overview system, I intend to use it.
	BI_2	I plan to use the assessment overview system in the future.
	BI_3	Given that I had access to the assessment overview system, I predict that I would use it.

Note: * - inverted item

5.3.3. Data collection procedure

Before the data collection, an ethical authorisation was obtained from the Ethics Committee at the university as well as at the industrial partner institutions. All expected ethical procedures were followed in the development and the administration of this study. The study was carried out with operator apprentices working in chemical industry and training institution situated in Germany and Belgium, respectively over a 2-month period (October 2021 to November 2021).

Apprentices were invited to voluntarily take part in the study. Prior to the start of the experiment, participants who agreed to take part were randomly designated to control group or experimental group and were asked to complete the pre-training questionnaire. Once they answered the pre-training questionnaire, the participants were given a brief introduction as well as a demonstration of the use of the Oculus Quest 2 device. After completing the introductory session, the participants were instructed to begin the *n*-BuLi production process in IVR environment. Upon the completion of the IVR training, participants were asked to complete the post-training questionnaire. In addition, a number of participants from both control group and experimental group were asked to complete the follow-up questionnaire.

After screening for incomplete and/or duplicated responses, a total of 32 and 31 respondents for control and experimental groups, respectively, completed pre-training questionnaires and post-training questionnaires. These were then retained for the data analysis. Moreover, a total of 18 and 12 completed follow-up questionnaires from both groups were collected for the data analysis.

5.3.4. Data analysis

According to Delice (2010) a minimum sample size of 30 in every subgroup is recommended to perform quantitative data analysis. As the total number of collected responses for both control group and experimental group was greater than 30, the quantitative data collected from the number scale questions were analysed using SPSS 25.0 (from IBM SPSS Statistics). Since this study aims to examine the significance of the difference between two interested groups in one target variable (*i.e.* attitude, satisfaction based on content, or behavioural intention), a statistical comparison analysis was employed (Puteh *et al.*, 2017).

For qualitative research, Vasileiou *et al.* (2018) stated that the minimum sample size of at least 12 in every subgroup is recommended to reach data saturation to ensure further gathering of new data no longer reveals new insights. Given that the total number of collected responses for both control group and experimental group was greater than 12, the qualitative data collected from the open-ended questions were analysed using inductive thematic analysis (Braun & Clarke, 2006; Wallace *et al.*, 2018). The inductive thematic analysis follows a six-step process: become familiar with the data, generate initial codes, search for potential themes, review/validate themes, further define themes, and report themes (Braun & Clarke, 2006; Wallace *et al.*, 2018). The first step of the process is to familiarize with the data which typically involves reading, examining, and re-examining the responses of the operator apprentices for each of the open-ended questions. Afterwards, create a list of preliminary codes (*e.g.* words/phrases) which highlight the vital information that are apparently relevant from the given data. Then, group the coded data with similar patterns into potential themes. After creation of themes, review and further refine them to avoid mismatch between the codes and the themes, as well as the themes with respect to the entire data set. Subsequently, determine the specific part of the given data every theme represents. Finally, create a visualisation report such as pie chart, bar chart, word cloud, as appropriate.

5.4. Results

5.4.1. Participant Profile

The demographic information of the participants is summarised in Table 5.2. The majority of the participants in both the control and the experimental group were male (78.13% and 80.65%, respectively) and between the ages of 20-30 (81.25% and 77.42%, respectively).

Table 5.2. Demographic information of participants.

Characteristics	**Items**	**Control group (n = 32)**		**Experimental group (n = 31)**	
		Frequency	**Percentage**	**Frequency**	**Percentage**
Gender	Male	25	78.13	25	80.65
	Female	4	12.50	6	19.35
	No answer / Prefer not to say	3	9.38	0	0.00
Age	18-19	3	9.38	6	19.35
	20-30	26	81.25	24	77.42
	31-40	1	3.13	1	3.23
	No answer / Prefer not to say	2	6.25	0	0.00

5.4.2. Assessment of the quantitative data

Before conducting a statistical comparison analysis, it is important to analyse the reliability and the validity of the data in order to determine the consistency as well as the accuracy of the questionnaire. To test the reliability, Cronbach's alpha was calculated. The calculated values of reliability and convergent validity for the control group and the experimental group is shown in Table 5.3. The Cronbach's alpha values for all factors were above the minimum cut-off of 0.6 (Nawi *et al.*, 2020), which indicates that the constructs have a strong reliability for each group.

Table 5.3. Reliability and convergent validity analysis of participants.

Constructs	Items	Factor Loading	Cronbach's Alpha	AVE[a]
Attitude	ATT_1	0.917 (0.855)	0.816 (0.738)	0.793 (0.768)
	ATT_2	0.863 (0.897)		
	ATT_3*			
	ATT_4*			
Satisfaction based on content	Sat_1	0.947 (0.893)	0.938 (0.904)	0.869 (0.827)
	Sat_2	0.915 (0.938)		
	Sat_3	0.934 (0.897)		
Behavioural Intention	BI_1	0.880 (0.762)	0.790 (0.751)	0.808 (0.746)
	BI_2	0.918 (0.954)		
	BI_3*			

Note: * - Removed due to the lack of outer loading reliability (< 0.7)
[a] - Average Variance Extracted
Numbers in bracket - Values for experimental group

In order to test the validity, the convergent and discriminant validities were calculated. Convergent validity (*i.e.* degree to which a given measure positively correlates with alternative measures of the same construct) is considered satisfactory if the value of the average variance extracted (AVE) and the factor loading is greater than 0.5 and 0.708, respectively (Hair *et al.*, 2017). As shown in Table 5.3, all constructs in every group had an AVE and factor loading values higher than the minimum cut-off. Furthermore, the Heterotrait-Monotrait ratio (HTMT) was calculated to evaluate the discriminant validity (*i.e.* the extent to which the items within a construct are truly distinct from each other) (Hair *et al.*, 2017). The values of discriminant validity for both groups are shown in Table 5.4 and they were lower than the threshold value of 0.9 for all combination of constructs (Gold *et al.*, 2001). Thus, the results obtained indicate adequate convergent and discriminant validities for each group.

Table 5.4. Discriminant validity analysis using Heterotrait-Monotrait (HTMT) ratio of participants.

	Attitude	Behavioural Intention	Satisfaction based on content
Attitude			
Behavioural Intention	0.354 (0.414)		
Satisfaction based on content	0.383 (0.220)	0.043 (0.283)	

Note: Numbers in bracket - Values for experimental group

Although it is important to establish reliability and validity of each group, it is also essential to analyse the normality of the data as this will dictate the type of a statistical comparison test to be used for this study, whether to use parametric *t*-test or a non-parametric Mann-Whitney U test. To determine whether the data is parametric or non-parametric, a Shapiro-Wilk test was conducted. Table 5.5 shows the summary of the Shapiro-Wilk test for the control group and the experimental group. Only the satisfaction based on content construct in the control group passed the normality test. Given that the distribution of data for other constructs in every group was not normal, the Mann-Whitney U test was used to test H1, H2, and H3.

Table 5.5. Summary of the normality test.

Constructs	Groups	Test of normality Shapiro-Wilk P value*	Result
Attitude	Control	0.003	Not normal
	Experimental	0.019	Not normal
Satisfaction based on content	Control	0.179	Normal
	Experimental	0.026	Not normal
Behavioural Intention	Control	0.050	Not normal
	Experimental	0.000	Not normal

Note: * = *p*-value greater than 0.05 indicates normal distribution

Having evaluated the reliability, validity, and normality of the data, Mann-Whitney U, a non-parametric test, was used to assess the similarities and the differences of mean ranks scores of target construct (attitude, satisfaction based on content, and behavioural intention) between the control group and experimental group. In this method, if the *p*-value is lower than 0.05, there is a significant difference in the target construct between two subpopulations. The outcome of the Mann-Whitney U test *p*-values in Table 5.6 showed that there were no significant differences in attitude, satisfaction based on content, and behavioural intention between the considered groups. Therefore, H1, H2, and H3 were not accepted.

Table 5.6. Outcomes of Mann-Whitney U test.

Constructs	**Groups**	**Mean**	**Mean rank**	**Sum of ranks**	**Mann-Whitney U**	***Z***	***P***
Attitude	Control	4.656	32.61	1043.50	476.5	-0.273	0.785
	Experimental	4.500	31.37	972.50			
Satisfaction based on content	Control	4.031	27.81	890.00	362.0	-1.858	0.063
	Experimental	4.548	36.32	1126.00			
Behavioural Intention	Control	4.219	30.33	970.50	442.5	-0.759	0.448
	Experimental	4.419	33.73	1045.50			

5.4.3. Assessment of qualitative data

In order to further investigate and evaluate the perspective of the participants with regards to the delivery and the presentation of their performance evaluation in VR-based training, two open-ended questions were posed to the participants in both groups. They explored the advantages and the disadvantages of the assessment system, respectively.

Seven themes (no advantages, no opinion, provides immediate results, helpful, great, additional motivation, and simplicity) emerged from the question 'What do you think are the advantages of using the assessment overview system?' for the control group while four themes (no opinion, provides immediate results, great, and additional motivation) for the experimental group (Figure 5.3).

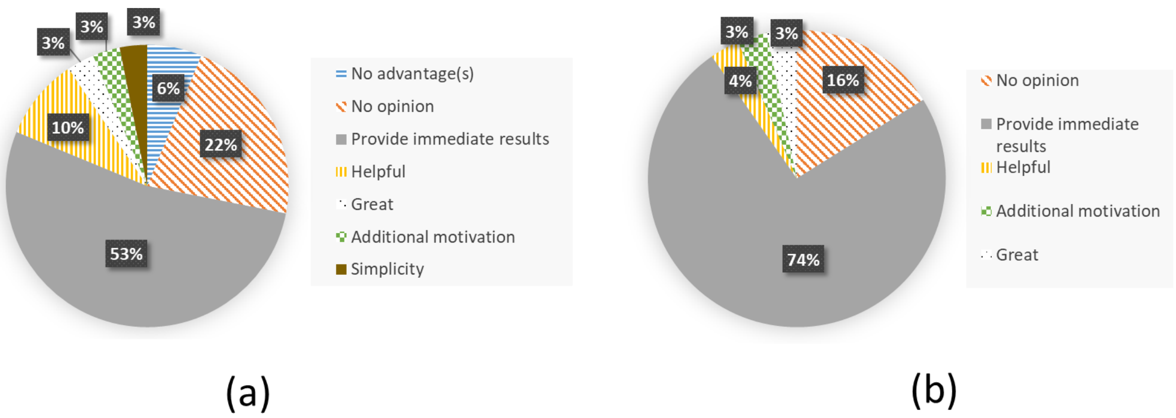


Figure 5.3. Advantages of (a) simplified assessment overview system and (b) detailed assessment overview system.

On the other hand, four themes (no disadvantage, no opinion, does not provide good feedback, and cause of additional stress) emerged from the question 'What do you think are the disadvantages of using the assessment overview system?' for both the control group and the experimental group (Figure 5.4).

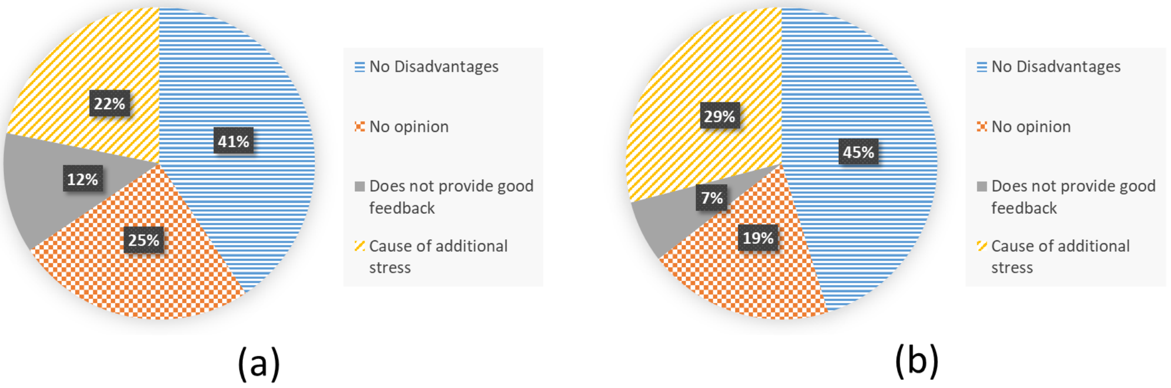


Figure 5.4. Disadvantages of (a) simplified assessment overview system and (b) detailed assessment overview system.

In addition, a number of participants from both groups were asked follow-up questions to examine the preference with regards to the presentation of their performance rating in the IVR environment. As shown in Figure 5.5, regardless of what kind of assessment overview system they were initially presented, majority of the participants from both groups prefer detailed assessment overview system over the simple assessment overview system.

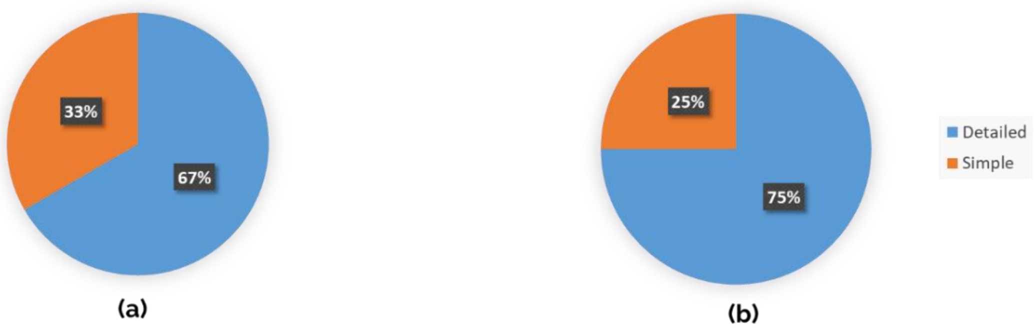


Figure 5.5. Assessment system preference for (a) control group participants and (b) experimental group participants.

The main reason why most of the chemical operator apprentices from both groups prefer detailed assessment overview system is because this assessment system has the ability to provide better feedback which allows them to see which area(s) need to be improved. Below are the responses of the chemical operator apprentices that highlight this claim from the control group:

‘You know more precisely where the error was.’

‘It is easier to recognize the mistakes made and to draw conclusions to learn from them.’

‘You can better see where the mistakes were.’

‘Because I can see everything where my error is.’

‘So you can see where the mistakes lie.’

Moreover, the responses of the chemical operator apprentices from the experimental group emphasise the abovementioned claim, as shown below:

‘Here (detailed assessment overview system), you get more specific information. So that you know where you are struggling.’

‘You can see what you can improve on.’

‘This way you can really know which parts need to be improved.’

‘You can see more details about what you have performed.’

‘More achievements to see and a better picture of how you did it.’

5.5. Discussion

The current study applied a *quasi*-experimental approach to examine the preferences of chemical operator apprentices towards how assessment should be presented in an IVR-based chemical production training called ‘Operate your own reactor’. This was done by comparing the mean rank scores on user attitude, satisfaction based on content, and on behavioural intention constructs, and by analysing the possible patterns in open text responses of the participants. Upon comparison of the groups of chemical operator apprentices, based on the presentation of the assessment system using Mann-Whitney U test and inductive thematic analysis, there were several similarities and differences in the constructs investigated in the current study.

5.5.1. Theoretical implications

This study provides meaningful insights on how the results of assessment should be presented within a VR setting for formative assessment. The empirical results showed that the mean values of all the considered constructs (Table 5.6) for both the control group (users who received simplified version of their performance results) and the experimental group (users who received detailed version of their performance results) were above 4 and ranging from 4.031 to 4.656 (mean value of 4 or higher indicates positive rating). This suggests that regardless of the group, the majority of the participants expressed a positive attitude, satisfaction based on content, and behavioural intention on assessment system. The positive mean values of all constructs for both groups are also reflected in their open text responses. For instance, participants from both groups have mentioned several advantages of using the assessment results overview system such as ‘provides immediate results’, ‘helpful’, ‘great’ and ‘additional motivation’ but only mentioned two distinct disadvantages (*e.g.* causes additional stress and does not provide good feedback). Given that they provided more advantages than disadvantages for both groups, this indicates that most of the participants demonstrated positive attitudes, positive satisfactions based on content, and positive behavioural intentions on the assessment system.

The questionnaire was designed to measure both the affective and cognitive components of attitude. Consequently, in the hypothesis testing, the positive scores of attitudes on both assessments (simplified and detailed) are consistent with the previous studies, such as Atek *et al.* (2012) and Andrew *et al.* (2018), that confirmed that users are more likely to have highly positive attitude scores towards a system that corresponds to their affective and cognitive impressions. This suggests that regardless of the comprehensiveness of the presentation of the OYOR assessment system, chemical operator apprentices express positive attitude toward the given assessment overview system as long as it can provide the needed performance evaluation score to pass the training.

Moreover, the positive mean values of satisfaction based on content on both assessment systems are consistent with the findings by Fitriantoro and Husnah (2018) that when the users know that the learning platform will provide the necessary information, they feel contented on that system. This observation suggests that the apprentices from both the control and the experimental group are contented with the amount of information provided by the respective OYOR assessment system.

Furthermore, regardless of the presentation scheme, participants from both groups believe that the use of OYOR assessment system is beneficial in terms of understanding their respective performance score. Thus, they gave positive rating for the behavioural intention construct. These positive ratings of behavioural intention on both assessment systems are also consistent with the findings by Udeozor *et al.* (2021) that the likelihood of the learner using a new system in the future will increase if such method is shown to significantly help the learner in achieving his/her objectives.

However, the results of the Mann-Whitney U test revealed that there is no significant group difference in their attitude, satisfaction based on content, and behavioural intention (Table 5.6). The probable rationale why there is no significant group difference for the three constructs discussed is that the users are mainly concerned with the assessment system's primary purpose which is to ultimately provide performance evaluation.

The qualitative data showed that regardless of the assessment system, participants perceived these tools as beneficial since they can provide immediate results presented after the completion of the training. This supports the results of the quantitative analysis that shows no significant group difference on attitude, satisfaction, and behavioural intention. On the other hand, more than 40% of the participants in both the control group and the experimental group stated that there were no disadvantages of using the simple assessment system or the detailed assessment system, respectively. This observation also suggests the idea that the ability of the assessment system to provide immediate results presented after the completion of the training is the primary concern for the chemical operator apprentices. Moreover, 25% and 19% of the participants in the control and the experimental group did not provide any comments with respect to the disadvantage(s) of the assessment system.

Furthermore, 12% of the participants in the control group stated that the simplified assessment system does not provide good feedback. Some stated that they were not informed of what they did correctly or incorrectly and thus the feedback is not helpful in further improving their skills. On the other hand, 7% of the participants in the experimental group pointed out that the detailed assessment system does not provide good feedback. Some participants from the experimental group stated that they wanted to receive more feedback as the system did not specify detailed feedback on areas where they performed poorly. Although these numbers were relatively low, this implies that some users wanted to know more about their actions to learn from them.

Lastly, 22% and 29% of the participants from the control and experimental groups, respectively, stated that the assessment system was not good as showing grades causes additional stress. For instance, some of the respondents indicated that:

'In case of poor performance: stress, and bad mood.'

'You get demotivated when you get a bad rating more often.'

'Maybe if you don't perform at a good level it's discouraging to see the scores every time.'

‘If you don't perform at a good level it might come across a bit harsh.’

According to Lerner and Tetlock (1999), performance evaluation is stressful as this causes trainees to feel some extent of pressure to fulfil the desired outcome. Moreover, Yılmaz *et al.* (2020) argued that some of the students who received low scores from the automated digital formative assessment feel demotivated. In this study, some respondents specifically said that they get demotivated from receiving a bad rating or that it causes them stress and puts them into a bad mood (refer to the responses above). Thus, some participants from both groups indicate that they expect some extent of additional stress from the presentation of assessment results regardless of the amount of information contained therein.

The novelty of the OYOR system to the chemical operator apprentices may have contributed to their contentment of whatever assessment presentation system have produced. This might be another reason why there is no significant group difference in their attitude, satisfaction based on content, and behavioural intention. In order to check if the introduction of new assessment overview system would result in a preference, a number of participants from both groups were asked to choose their preference with regards to the presentation of their performance rating in the IVR environment.

With respect to the abovementioned question, the majority of the participants from both groups prefer detailed assessment overview system over the simple assessment overview system regardless of what kind of assessment overview system they were initially presented within the OYOR prototype (Figure 5.5). This result is also consistent with previous studies that confirmed the preference of the users towards detailed assessment in online computer-based testing (Baleni, 2015; Bulut *et al.*, 2019; Chin *et al.*, 2021). The main reason for the apprentices who chose detailed assessment system is the fact that it provides better feedback which allows them to see which area(s) need to be improved (for the detailed reasons of the respondents, refer to section 5.4.3). This indicates that when users know that the given assessment system provides these attractive learning features such as good and useful feedback to improve their individual performance, they tend to prefer that assessment system (Baleni, 2015).

5.5.2. Practical implications

The findings of this study have a number of practical implications in employing assessment in a long procedural IVR-based training. If institutions decide to develop and implement an assessment system in IVR environment, they should focus more on enhancing the quality of the immediate performance results and/or feedback. This is because the ability of the assessment system to provide and present immediate performance results and/or feedback is the most important factor for the majority of the chemical operator apprentices from both the control group and the experimental group. This study will aid assessment designers in having a clearer view on what aspects they need to focus on more in order to design an effective assessment system which can better support independent learning through continuous provision of feedback. Although there were no significant differences in the perspective of the participants with regards to the delivery and the presentation of their performance evaluation in VR-based training (simple vs. detailed), it was evident that most of the chemical operator apprentices were in favour of using the detailed assessment system in the OYOR prototype. As elaborated by some of the responses of the apprentices in the previous sub-chapter, the detailed assessment system provides specific performance results which can be used for their self-improvement. Thus, the current study recommends that assessment designers give great consideration in employing the concept of detailed assessment system in their IVR-based training projects.

5.6. Conclusions

The present study uses both Mann-Whitney U test (quantitative analysis) and inductive thematic analysis (qualitative analysis) to examine the views of 63 chemical operator apprentices toward how assessment should be presented in ‘Operate your own reactor’ prototype. The results of the quantitative analysis showed that there were no significant group differences between the control group of users (who received a simplified version of their performance results) and the experimental group of users (who received a detailed version of their performance results) in terms of their attitude, satisfaction based on content, and behavioural intention in IVR training environment. Moreover, when determining this, it was evident from the results of the qualitative analysis that the users are mainly concerned with the primary purpose of the assessment system to ultimately provide immediate performance evaluation results presented after the completion of the training. This is the reason why both groups demonstrated positive attitude, satisfaction based on content, and behavioural intention regardless of the differences in the assessment presentation. However, in terms of the user preference, most of the participants in both groups prefer detailed assessment results as they can review and improve their performance for the future. In summary, this study provides useful insights for assessment designers in formulating suitable schemes on the optimal ways in presenting and delivering assessment results in a long procedural IVR-based training.

5.7. Note

Part of this chapter has been submitted (September 2022) in the Education and Information Technologies Journal: Toyoda, R., Garcia-Fracaro, S., Gallagher, T., Tehreem, Y., Russo Abegão, F., & Glassey, J. Incorporating assessment into chemical industry training through immersive virtual reality (IVR): a case study

Chapter 6. Conclusions and recommendations for future work

6.1. Conclusions

Industries with inherent high operational risks, such as chemical plants, must put in place mitigation measures to avoid accidents and guarantee the safety of its employees. One of these measures is to ensure that the operators are equipped with proper and effective training that is primarily designed to achieve mastery of their respective functions. In order to achieve this objective, the training environment must be engaging and functional, a systematic determination of the expertise level of the trainees must be established, the assessment framework in evaluating the training outcome must be properly laid down, and such system must be proven and/or tested.

It is in this context that the Virtual Reality-based training was introduced in chapter 2 of this thesis as a potential solution to address the need for a more effective training environment for chemical plant operators. This thesis chapter includes review of existing literature, consisting of 59 articles, for VR-based health and safety training in high-risk industries. These articles were analysed to collect information on the type of VRs used in training, topics of the H&S training, the assessment techniques implemented, among others. From this review, a comparison was made of IVR-based H&S training *vis-à-vis* traditional and/or other VR-based training method in terms of improving the training evaluation outcome.

The review of literature in connection to VR-based training suggests that most industries included in the study use such training environments to train users for risk assessment, machinery and/or equipment process/procedural operation, or both topics. It also showed the significant increase in the use of fully immersive VR in various industries due to its hardware improvements, enhanced display resolution, and a more competitive pricing. Moreover, existing literature indicates that stakeholders measure the effectiveness of training in terms of the learning achievements within a short period.

To evaluate the satisfaction with the VR-based H&S training, most researchers used external assessment methods, such as questionnaires and interviews, due to ease of their implementation. To evaluate the knowledge gain of trainees, researchers used knowledge tests and multiple-choice questionnaires. While most researchers implemented external assessment methods, some researchers used internal assessment techniques. These include the use of log data to develop an automated assessment system which has the capability to measure complex skills. The review of existing literature also collectively suggests that VR-based H&S training, in comparison to traditional training methods, has more potential to improve the reaction, learning, and behaviour levels of the trainees.

The results of Chapter 2 provide a basis for future research to consider the effectiveness of VR-based H&S training, measured using internal assessment methods, in behavioural level and organisational outcome level (third and fourth level of Kirkpatrick's model) of the trainees.

There was a noticeable increase in the popularity of VR-based tools, especially for immersive virtual reality (IVR), during the height of the COVID-19 pandemic. Despite this, some people still show hesitancy in adopting these tools. To resolve this, developers must acknowledge and understand the gap between what people claim, as manifested in their attitude and involvement, and how they behave. By understanding the interrelationship between behavioural intention and influential factors, and how to harmonize them, developers will be able to improve the rate of success in IVR training implementation. To gather useful insights on the perception of potential trainees and their behavioural intention towards adopting IVR-based training, Chapter 3 analysed data from 438 individuals using PLS-SEM and multi-group analysis (MGA) with SmartPLS 3.0 software. Data on prior-immersive VR-experience, nationality, and the length of work experience were used to group the respondents. The results of the study show no statistically significant differences among the predefined groups of respondents. Nonetheless, the MGA methodology may be effective in discerning intentions of various groups, which may be achieved by evaluating the ranking order of the constructs from the UTAUT model across group-specific results. Policymakers and stakeholders in high-risk industries, including chemical industry, may use the insights from this chapter in order to formulate effective strategies in implementing IVR-based technology in measured groups.

Aside from the formulation of the effective strategies such as whether to emphasize game elements, easier controls, more procedural aspects, *etc.*, there is a growing demand from stakeholders for the development of an unobtrusive and systematic assessment framework to further maximise the potential of IVR as a training environment. Hence, Chapter 4 proposed and developed, for the first time, a detailed methodological framework combining proven approaches, such as the Dreyfus model, evidence-centred design (ECD), and fuzzy comprehensive evaluation method, for the evaluation of user expertise in IVR-based training in chemical plant environment. Given that the assessment framework must determine the expertise level of the operators by thoroughly evaluating their training results, Dreyfus model was incorporated into the proposed IVR-based assessment framework. The Dreyfus model characterizes the skill growth of an individual from novice to advanced beginner, to competent, to proficient, and finally, to an expert.

The use of IVR may provide large volumes of data relating to the conduct and outcome of the training. To determine which aspects of these data is useful in evaluating the expertise of the trainees, the evidence-centred design (ECD) was incorporated in the framework. The ECD creates a logical relation between the collected data from IVR to the corresponding expertise level of the trainees. Finally, to ensure the precision of the expertise evaluation, the fuzzy comprehensive evaluation method was used to further delineate the expertise levels by enhancing the resolution of the measures in the Dreyfus model.

To test the assessment framework, an IVR-training called 'Operate Your Own Reactor' (OYOR) for 19 operator apprentices in ACTA (Belgium) was conducted. Data from the IVR training, such as number of mistakes committed, total time of completion, and number of hints requested were gathered. Guided by the ECD framework, which was further enhanced by Fuzzy evaluation method, the expertise levels of all training participants, in accordance with the Dreyfus model, were determined.

The integration of the proposed assessment framework in OYOR environment has broad implications for various stakeholders. For instance, trainees can use the summary of the user performance result in the OYOR environment (Figure 4.7) to determine their strengths and weaknesses in the performance of their tasks. Moreover, trainers and administrators can use the visualisation reporting tool (Figure 4.8-4.13) to provide more constructive and personalised feedback for trainees to ensure their continual growth and to expedite the development of chemical reactor operator experts leading to more cost-effective and efficient industry operation, respectively.

To provide a holistic approach for the improvement and optimization of IVR-based H&S training, chapter 5 determined the probable overall preference of a user in terms of the manner on how assessment results are presented. The responses of 63 operator apprentices, using the OYOR prototype, were analysed using various methods. The respondents were first grouped into two groups, where one group was presented with simplified training results while the other group was presented with detailed assessment results. In terms of attitude, satisfaction based on content, and behavioural intention, the two groups did not show statistically significant differences based on Man-Whitney U test. From the inductive thematic analysis, it was observed that users are mainly interested in immediately acquiring their performance evaluation results. Hence, both groups showed positive attitude, satisfaction based on content, and behavioural intention. However, in terms of user preference, respondents from both groups showed preference for a detailed assessment result. This is because they are given the option to comprehensively review their past performance and improve thereon. Training developers and assessment designers may consider these findings to formulate effective and suitable manners of presenting assessment results, especially for a more complex IVR-based training.

6.2. Recommendations for future work

Although this study provides a deeper understanding of VR-based H&S training for researchers and stakeholders, it is important to acknowledge some limitations. To start with, the literature review section (Chapter 2) only explored VR-based H&S training applications for high-risk industries. It is worth acknowledging that the validity of the conclusions from this study is within the scope of aforementioned research boundaries. To further improve on this research, the application of the VR-based training may be expanded to its potential integration in different workplace training, such as communication, and leadership training to visualise the whole picture on VR-based workplace training and to establish the wider applicability of the results. Moreover, as the literature review section only considered peer-reviewed articles from the largest digital source available (Scopus) during literature search, it may be beneficial to consider other type of articles (*e.g.* professional reports, research project deliverables, trade publications) from other databases as there may be interesting results from these type of articles.

Secondly, the responses from the IVR-based technology adoption intention study (Chapter 3) were obtained from convenience sampling as most of the chemical industries do not openly publish the details of their employees to ensure compliance with privacy regulations. For the future studies, it would be beneficial to replicate this study using different groups from a wider and a more heterogeneous population (*e.g.* chemical industries from different parts of the world and different sectors) to establish the robustness of the results. Moreover, as this study only used a quantitative statistical approach to examine the relationship between factors, the exploration of qualitative approaches or a combination of both methods for better and more comprehensive explanations and perceptions may be beneficial to further investigate the mechanism of behavioural intention among this group of people. Furthermore, since the perception of users with respect to IVR adoption in chemical industry may change over time, longitudinal studies at a various timeline of IVR acceptance process to reinvestigate the IVR adoption considering other key constructs may be valuable.

Thirdly, although the newly developed and implemented assessment framework will benefit stakeholders and administrators in producing experts, it is important to take note that all industries and organisations may have specific sets of standards in their processes which may render the assessment criteria and their corresponding descriptors used herein not entirely applicable. This will require re-validation and modification of such criteria to fit particular industry standards. Moreover, as the formulation of assessment rubrics requires validation by experts, the availability of experts in the field may also represent some limitations in the verification process. If there are sufficient available experts in the field, it may be beneficial for the future studies to consider other methodologies in arriving at a consensus of the experts. Furthermore, an additional sub-process should be further conducted to properly assign weights to specific assessment rubrics, for the purposes of determining the overall performance of the group or the organisation. For instance, the most critical criteria (*e.g.* number of mistakes or even a seriousness of specific mistakes, if such a measure is defined in a given application) could be assigned a higher weighting in comparison to other criteria which pertain to more innocuous measures (*e.g.* number of hints received). Through assigning more appropriate weights, a trainee who committed a substantial error or underperformed in critical criteria will most likely have poor overall evaluation. On the other hand, trainees who committed a number of errors in less serious criteria may still obtain an overall good evaluation if the cumulative effect of less serious errors does not lead to a major failure. The more refined weighted approach can ensure a more comprehensive and holistic calculation of the overall scores vis-à-vis a specific assessment criterion.

Finally, although the fifth chapter used both quantitative and qualitative methods to examine the preference of the users on assessment results system, the interpretation was based on a single population (chemical operator apprentices). As every individual has their own predispositions and opinions on the implementation and the use of digital assessment, it would be beneficial for any future study to replicate this study using different groups using a more comprehensive population, such as experienced operators, university students, trainers, *etc.*, to achieve deeper insights into issues related to designing, administering, and presenting assessment in IVR environment. Likewise, since the perception of users with respect to the delivery and presentation of their performance in IVR environment may change over time, longitudinal studies would also be valuable to identify whether there is a relationship between their frequency of usage and their preference toward assessment presentation. Furthermore, a future study may be conducted to measure the degree of the stress or demotivation felt by respondents in receiving a generally unfavourable assessment results and its correlation to how such results were presented.

References

Adami, P., Rodrigues, P. B., Woods, P. J., Becerik-Gerber, B., Soibelman, L., Copur-Gencturk, Y., & Lucas, G. (2021). Effectiveness of VR-based training on improving construction workers' knowledge, skills, and safety behavior in robotic teleoperation. *Advanced Engineering Informatics*, *50*. https://doi.org/10.1016/j.aei.2021.101431

Ahmad, W. F. W., Sarlan, A., & Rauf, N. A. (2019). *ESCAPE: Interactive Fire Simulation and Training for Children Using Virtual Reality* (Vol. 843, Issue January, pp. 670–679). Springer International Publishing. https://doi.org/10.1007/978-3-319-99007-1_62

Ahn, S., Kim, T., Park, Y.-J., & Kim, J.-M. (2020). Improving Effectiveness of Safety Training at Construction Worksite Using 3D BIM Simulation. *Advances in Civil Engineering*, *2020*, 1–12. https://doi.org/10.1155/2020/2473138

Albers, A., Turki, T., & Lohmeyer, Q. (2012). Assessment of design competencies by a five level model of expertise. *14th International Conference on Engineering and Product Design Education (E&PDE)*.

Albert, A., Hallowell, M. R., Kleiner, B., Chen, A., & Golparvar-Fard, M. (2014). Enhancing Construction Hazard Recognition with High-Fidelity Augmented Virtuality. *Journal of Construction Engineering and Management*, *140*(7), 04014024. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000860

Alcañiz, M., Parra, E., & Giglioli, I. A. C. (2018). Virtual reality as an emerging methodology for leadership assessment and training. *Frontiers in Psychology*, *9*(SEP), 1–7. https://doi.org/10.3389/fpsyg.2018.01658

Alexander, P. A. (2003). The Development of Expertise: The Journey From Acclimation to Proficiency. *Educational Researcher*, *32*(8), 10–14. https://doi.org/10.3102/0013189X032008010

Alexander, P. A. (2004). A model of domain learning: Reinterpreting expertise as a multidimensional, multistage process. In *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development.* (pp. 273–298). Lawrence Erlbaum Associates Publishers.

Allen, D., & Tanner, K. (2006). Rubrics: Tools for Making Learning Goals and Evaluation Criteria Explicit for Both Teachers and Learners. *CBE—Life Sciences Education*, *5*(3), 197–203. https://doi.org/10.1187/cbe.06-06-0168

Alraja, M. N. (2016). The Effect of Social Influence and Facilitating Conditions on E-government Acceptance from the Individual Employees' Perspective. *Polish Journal of*

Management Studies, *14*(2), 18–27. https://doi.org/10.17512/pjms.2016.14.2.02

Alsalamah, A., & Callinan, C. (2021). The Kirkpatrick model for training evaluation: bibliometric analysis after 60 years (1959–2020). In *Industrial and Commercial Training* (Vol. 54, Issue 1, pp. 36–63). Emerald Group Holdings Ltd. https://doi.org/10.1108/ICT-12-2020-0115

An, C. M., & Park, Y. H. (2018). The effects of semi-immersive virtual reality therapy on standing balance and upright mobility function in individuals with chronic incomplete spinal cord injury: A preliminary study. *Journal of Spinal Cord Medicine*, *41*(2), 223–229. https://doi.org/10.1080/10790268.2017.1369217

Andrew, M., Taylorson, J., Langille, D. J., Grange, A., & Williams, N. (2018). Student attitudes towards technology and their preferences for learning tools/devices at two universities in the UAE. *Journal of Information Technology Education: Research*, *17*, 309–344. https://doi.org/10.28945/4111

Arkorful, V., & Abaidoo, N. (2015). The role of e-learning, advantages and disadvantages of its adoption in higher education. *International Journal of Instructional Technology and Distance Learning*, *12*(1), 29–42.

Atek, E. S. E., Jusoh, Z., Alias, A. N., Wahid, W. A., & Tahir, I. M. (2012). Students' Attitudes towards the English Proficiency Enhancement Programme. *International Journal of Education*, *4*(3). https://doi.org/10.5296/ije.v4i3.1475

Attri, R. K. (2019). *Models Of Skill Acquisition And Expertise Development : A Quick Reference Of Summaries.* Speed To Proficiency Research: S2pro(C).

Avveduto, G., Tanca, C., Lorenzini, C., Tecchia, F., Carrozzino, M., & Bergamasco, M. (2017). Safety Training Using Virtual Reality: A Comparative Approach. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 10324 LNCS* (pp. 148–163). https://doi.org/10.1007/978-3-319-60922-5_11

Ayala García, A., Galván Bobadilla, I., Arroyo Figueroa, G., Pérez Ramírez, M., & Muñoz Román, J. (2016). Virtual reality training system for maintenance and operation of high-voltage overhead power lines. *Virtual Reality*, *20*(1), 27–40. https://doi.org/10.1007/s10055-015-0280-6

Baleni, Z. G. (2015). Online formative assessment in higher education: Its pros and cons. *Electronic Journal of E-Learning* , *13*(4), 228–236.

Barana, A., Marchisio, M., & Sacchet, M. (2019). Advantages of using automatic formative assessment for learning mathematics. *Communications in Computer and Information Science*, *1014*, 180–198. https://doi.org/10.1007/978-3-030-25264-9_12

Barra, E., López-Pernas, S., Alonso, A., Sánchez-Rada, J. F., Gordillo, A., & Quemada, J. (2020). Automated Assessment in Programming Courses: A Case Study during the COVID-19 Era. *Sustainability (Switzerland)*, *12*(18). https://doi.org/10.3390/SU12187451

Bates, R. (2004). A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, *27*(3), 341–347. https://doi.org/10.1016/j.evalprogplan.2004.04.011

Beh, H. J., Rashidi, A., Talei, A., & Lee, Y. S. (2021). Developing engineering students' capabilities through game-based virtual reality technology for building utility inspection. *Engineering, Construction and Architectural Management*. https://doi.org/10.1108/ECAM-02-2021-0174

Benner, P. (1982). From Novice to Expert. *American Journal of Nursing*, *82*(3), 402–407.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy and Practice*, *18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Bernard, F., Bonnardel, X., Paquin, R., Petit, M., Marandel, K., Bordin, N., & Bonnardel, F. (2021). Digital simulation tools in aviation maintainability training. *Computer Applications in Engineering Education*. https://doi.org/10.1002/cae.22461

Beyer-Berjot, L., Patel, V., Acharya, A., Taylor, D., Bonrath, E., Grantcharov, T., Darzi, A., & Aggarwal, R. (2014). Surgical training: Design of a virtual care pathway approach. *Surgery (United States)*, *156*(3), 689–697. https://doi.org/10.1016/j.surg.2014.04.045

Bissonnette, V., Mirchi, N., Ledwos, N., Alsidieri, G., Winkler-Schwartz, A., & Del Maestro, R. F. (2019). Artificial Intelligence Distinguishes Surgical Training Levels in a Virtual Reality Spinal Task. *The Journal of Bone and Joint Surgery*, *101*(23), e127. https://doi.org/10.2106/JBJS.18.01197

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *International Journal of Phytoremediation*, *21*(1), 7–74. https://doi.org/10.1080/0969595980050102

Blickensderfer, E. L., Strally, S., & Doherty, S. (2012). The Effects of Scenario-Based Training on Pilots' Use of an Emergency Whole-Plane Parachute. *International Journal of Aviation Psychology*, *22*(2), 184–202. https://doi.org/10.1080/10508414.2012.663253

Bozan, K., Parker, K., & Davey, B. (2016). A closer look at the social influence construct in the UTAUT Model: An institutional theory based approach to investigate health IT adoption patterns of the elderly. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2016-March*, 3105–3114. https://doi.org/10.1109/HICSS.2016.391

Bracq, M.-S., Michinov, E., Arnaldi, B., Caillaud, B., Gibaud, B., Gouranton, V., & Jannin, P. (2019). Learning procedural skills with a virtual reality simulator: An acceptability study. *Nurse Education Today*, *79*(February), 153–160. https://doi.org/10.1016/j.nedt.2019.05.026

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brecher, R. W. (1997). Risk Assessment. *Toxicologic Pathology*, *25*(1), 23–26. https://doi.org/10.1177/019262339702500106

Brookhart, S. M. (1999). *The Art and Science of Classroom Assessment. The Missing Part of Pedagogy. ASHE-ERIC Higher Education Report, Volume 27, Number 1.* ERIC.

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD.

Bulut, O., Cutumisu, M., Aquilina, A. M., & Singh, D. (2019). Effects of Digital Score Reporting and Feedback on Students' Learning in Higher Education. *Frontiers in Education*, *4*. https://doi.org/10.3389/feduc.2019.00065

Burdea, G. (2003). *Virtual reality technology* (2nd ed..). J. Wiley-Interscience.

Bushnell, D.S. (1990), "Input, process, output: a model for evaluating training", Training & Development Journal, Vol. 44No. 3, pp. 41-43.

Caruana, E. J., Roman, M., Hernández-Sánchez, J., & Solli, P. (2015). Longitudinal studies. *Journal of Thoracic Disease*, *7*(11), E537–E540. https://doi.org/10.3978/j.issn.2072-1439.2015.10.63

Chan, P., Van Gerven, T., Dubois, J. L., & Bernaerts, K. (2021). Design and Development of a VR Serious Game for Chemical Laboratory Safety. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13134 LNCS*, 23–33. https://doi.org/10.1007/978-3-030-92182-8_3

Cheah, J.-H., Thurasamy, R., Memon, M. A., Chuah, F., & Ting, H. (2020). Multigroup Analysis using SmartPLS: Step-by-Step Guidelines for Business Research. *Asian Journal of Business Research*, *10*(3), I–XIX. https://doi.org/10.14707/ajbr.200087

Checa, D., & Bustillo, A. (2019). A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications*, 5501–5527. https://doi.org/10.1007/s11042-019-08348-9

Chen, J. F., Hsieh, H. N., & Do, Q. H. (2015). Evaluating teaching performance based on fuzzy AHP and comprehensive evaluation approach. *Applied Soft Computing Journal*, *28*, 100–108. https://doi.org/10.1016/j.asoc.2014.11.050

Chiao, H. M., Chen, Y. L., & Huang, W. H. (2018). Examining the usability of an online virtual tour-guiding platform for cultural tourism education. *Journal of Hospitality, Leisure, Sport and Tourism Education*, *23*(August 2017), 29–38. https://doi.org/10.1016/j.jhlste.2018.05.002

Chin, H., Chew, C. M., & Lim, H. L. (2021). Incorporating feedback in online cognitive diagnostic assessment for enhancing grade five students' achievement in 'time.' *Journal of Computers in Education*, *8*(2), 183–212. https://doi.org/10.1007/s40692-020-00176-3

Chin, J., Dukes, R., & Gamson, W. (2009). Assessment in Simulation and Gaming. *Simulation & Gaming*, *40*(4), 553–568. https://doi.org/10.1177/1046878109332955

Choi, M., Ahn, S., & Seo, J. (2020). VR-Based investigation of forklift operator situation awareness for preventing collision accidents. *Accident Analysis & Prevention*, *136*, 105404. https://doi.org/10.1016/j.aap.2019.105404

Chomeya, R. (2010). Quality of Psychology Test Between Likert Scale 5 and 6 Points. *Journal of Social Sciences*, *6*(3), 399–403. https://doi.org/10.3844/jssp.2010.399.403

Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology*, *9*(NOV), 1–20. https://doi.org/10.3389/fpsyg.2018.02086

Clarke-Midura, J., & Dede, C. (2010). Assessment, technology, and change. *Journal of Research on Technology in Education*, *42*(3), 309–328. https://doi.org/10.1080/15391523.2010.10782553

Clifford, R. M. S., Jung, S., Hoerrnann, S., Billinqhurst, M., & Lindeman, R. W. (2019). Creating a stressful decision making environment for aerial firefighter training in virtual reality. *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*, 181–189. https://doi.org/10.1109/VR.2019.8797889

Colombo, S., & Golzio, L. (2016). The Plant Simulator as viable means to prevent and manage risk through competencies management: Experiment results. *Safety Science*, *84*, 46–56. https://doi.org/10.1016/j.ssci.2015.11.021

Colombo, S., Nazir, S., & Manca, D. (2014). Immersive Virtual Reality for Training and Decision Making: Preliminary Results of Experiments Performed With a Plant Simulator. *SPE Economics & Management*, *6*(04), 165–172. https://doi.org/10.2118/164993-PA

Cronbach, L. J. (1992). Four Psychological Bulletin Articles in Perspective. *Psychological Bulletin*, *112*(3), 389–392. https://doi.org/10.1037/0033-2909.112.3.389

Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin*, *74*(1), 68–80. https://doi.org/10.1037/h0029382

Cukurova, B. M., Mavrikis, M., & Luckin, R. (2017). Evidence-Centered Design and Its Application to Collaborative Problem Solving in Practice-based Learning Environments About Analytics for Learning (A4L). January.

Czarnek, G., Strojny, P., Strojny, A., & Richter, M. (2019). Assessing Engagement during Rescue Operation Simulated in Virtual Reality: A Psychophysiological Study. *International Journal of Human–Computer Interaction*, *0*(00), 1–13. https://doi.org/10.1080/10447318.2019.1655905

da Silva Júnior, J. N., Sousa Lima, M. A., Ávila Pimenta, A. T., Nunes, F. M., Monteiro, Á. C., de Sousa, U. S., Leite Júnior, A. J. M., Zampieri, D., Oliveira Alexandre, F. S., de Sousa, U. S., Pacioni, N. L., & Winum, J.-Y. (2021). Design, implementation, and evaluation of a game-based application for aiding chemical engineering and chemistry students to review the organic reactions. *Education for Chemical Engineers*, *34*, 106–114. https://doi.org/10.1016/j.ece.2020.11.007

Dado, M., Kotek, L., Hnilica, R., & Tůma, Z. (2018). The Application of Virtual Reality for Hazard Identification Training in the Context of Machinery Safety: A Preliminary Study. *Manufacturing Technology*, *18*(5), 732–736. https://doi.org/10.21062/ujep/168.2018/a/1213-2489/MT/18/5/732

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly: Management Information Systems*, *13*(3), 319–339. https://doi.org/10.2307/249008

Davis, S., Nesbitt, K., & Nalivaiko, E. (2014). Systematic review of cybersickness. ACM International Conference Proceeding Series, 02-03-December-2014. https://doi.org/10.1145/2677758.2677780

Dawson-Howe, K. M. (1995). Automatic submission and administration of programming assignments. *ACM SIGCSE Bulletin*, *27*(4), 51–53. https://doi.org/10.1145/216511.216539

Delice, A. (2010). The Sampling Issues in Quantitative Research. *Educational Sciences: Theory and Practice*, *10*(4), 2001–2018.

Dhalmahapatra, K., Maiti, J., & Krishna, O. B. (2021). Assessment of virtual reality based safety training simulator for electric overhead crane operations. *Safety Science*, *139*. https://doi.org/10.1016/j.ssci.2021.105241

Dholakiya, N. D., Ferjencik, M., Schofield, D., & Kubík, J. (2019). Virtual learning for safety, why not a smartphone? *Process Safety Progress*, *38*(2). https://doi.org/10.1002/prs.12005

Diego-Mas, J. A., Alcaide-Marzal, J., & Poveda-Bautista, R. (2020). Effects of Using

Immersive Media on the Effectiveness of Training to Prevent Ergonomics Risks. *International Journal of Environmental Research and Public Health*, *17*(7), 2592. https://doi.org/10.3390/ijerph17072592

Domínguez Alfaro, J. L., Gantois, S., Blattgerste, J., De Croon, R., Verbert, K., Pfeiffer, T., & Van Puyvelde, P. (2022). Mobile Augmented Reality Laboratory for Learning Acid–Base Titration. *Journal of Chemical Education*, *99*(2), 531–537. https://doi.org/10.1021/acs.jchemed.1c00894

Dorer, B. (2012). *Round 6 Translation Guidelines*.

Dorozhkin, D., Olasky, J., Jones, D. B., Schwaitzberg, S. D., Jones, S. B., Cao, C. G. L., Molina, M., Henriques, S., Wang, J., Flinn, J., & De, S. (2017). OR fire virtual training simulator: design and face validity. *Surgical Endoscopy*, *31*(9), 3527–3533. https://doi.org/10.1007/s00464-016-5379-7

Drey, T., Jansen, P., Fischbach, F., Frommel, J., & Rukzio, E. (2020, April 25). Towards progress assessment for adaptive hints in educational virtual reality games. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3334480.3382789

Dreyfus, S. E. (2004). The five-stage model of adult skill acquisition. *Bulletin of Science, Technology and Society*, *24*(3), 177–181. https://doi.org/10.1177/0270467604264992

Dreyfus, S. E., & Dreyfus, H. L. (1987). Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer. *IEEE Expert*, *2*(2), 110–111. https://doi.org/10.1109/MEX.1987.4307079

Dumal, P. A. A., Shanika, W. K. D., Pathinayake, S. A. D., & Sandanayake, T. C. (2017). Adaptive and automated online assessment evaluation system. *2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, 1–8. https://doi.org/10.1109/SKIMA.2017.8294135

Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. In *Sensors (Switzerland)* (Vol. 20, Issue 3). MDPI AG. https://doi.org/10.3390/s20030592

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, *82*(397), 171–185. https://doi.org/10.1080/01621459.1987.10478410

Elgohary, E., & Abdelazyz, R. (2020). The impact of employees' resistance to change on implementing e-government systems: An empirical study in Egypt. *The Electronic Journal of Information Systems in Developing Countries*, *86*(6). https://doi.org/10.1002/isd2.12139

Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (Eds.). (2006). *The*

Cambridge Handbook of Expertise and Expert Performance. Cambridge University Press. https://doi.org/10.1017/CBO9780511816796

Eseryel, D., Ifenthaler, D., & Ge, X. (2011). Alternative Assessment Strategies for Complex Problem Solving in Game-Based Learning Environments. In *Multiple Perspectives on Problem Solving and Learning in the Digital Age* (pp. 159–178). Springer New York. https://doi.org/10.1007/978-1-4419-7612-3_11

Fällman, D., Backman, A., & Holmlund, K. (1999). VR in Education : An Introduction to Multisensory Constructivist Learning Environments. :

Farrell, R. (2012). Reconsidering the Relationship Between Generic and Situated IL Reconsidering the Relationship Between Generic and Situated IL Approaches: The Dreyfus Model of Skill Acquisition in Formal Approaches: The Dreyfus Model of Skill Acquisition in Formal Informa. *Library Philosophy and Practice*, 1–16.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Feise, H. J., & Schaer, E. (2021). Mastering digitized chemical engineering. *Education for Chemical Engineers*, *34*, 78–86. https://doi.org/10.1016/j.ece.2020.11.011

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 and Education*, *127*(May), 252–266. https://doi.org/10.1016/j.compedu.2018.09.002

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behaviour: An introduction to theory and research* (Vol. 27).

Fitriantoro, M. J., & Husnah, N. (2018). The Implementation of the End-User Computing Satisfaction Model into SCeLE: A Study of the Undergraduate Program of the Accounting Department in Universitas Indonesia. *Proceedings of the 6th International Accounting Conference (IAC 2017)*. https://doi.org/10.2991/iac-17.2018.27

Freina, L., & Canessa, A. (2015). Immersive vs Desktop Virtual Reality in Game Based Learning. In R. Munkvold & L. Kolås (Eds.), *9th European Conference on Games Based Learning* (pp. 195–202). Academic Conferences and Publishing International Limited.

Gallagher, T., Slof, B., van der Schaaf, M., Toyoda, R., Tehreem, Y., Fracaro, S. G., & Kester, L. (2022). Comparison with Self vs Comparison with Others: The Influence of Learning Analytics Dashboard Design on Learner Dashboard Use. In GALA conference (pp. 11–21). https://doi.org/10.1007/978-3-031-22124-8_2

Gallegos-Nieto, E., Medellín-Castillo, H. I., González-Badillo, G., Lim, T., & Ritchie, J.

(2017). The analysis and evaluation of the influence of haptic-enabled virtual assembly training on real assembly performance. *The International Journal of Advanced Manufacturing Technology*, *89*(1–4), 581–598. https://doi.org/10.1007/s00170-016-9120-4

Gao, Y., Gonzalez, V. A., & Yiu, T. W. (2019). The effectiveness of traditional tools and computer-aided technologies for health and safety training in the construction sector: A systematic review. *Computers and Education*, *138*(May), 101–115. https://doi.org/10.1016/j.compedu.2019.05.003

Gao, Y., González, V. A., & Yiu, T. W. (2017). Serious Games vs. Traditional Tools in Construction Safety Training: A Review. *Lean and Computing in Construction Congress - Volume 1: Proceedings of the Joint Conference on Computing in Construction*, *July*, 653–660. https://doi.org/10.24928/JC3-2017/0070

Garcia Fracaro, S., Wilk, W., Glassey, J., & Bernaerts, K. (2020). Preliminary feasibility study of virtual reality application for training of chemical operator apprentices. Chemie Ingenieur Technik, 92(9), 1196–1197. https://doi.org/10.1002/cite.202055338

Garcia Fracaro, S., Chan, P., Gallagher, T., Tehreem, Y., Toyoda, R., Bernaerts, K., Glassey, J., Pfeiffer, T., Slof, B., Wachsmuth, S., & Wilk, M. (2021). Towards design guidelines for virtual reality training for the chemical industry. *Education for Chemical Engineers*, *36*, 12–23. https://doi.org/10.1016/j.ece.2021.01.014

Ghaderi, I., Korovin, L., & Farrell, T. M. (2021). Preparation for Challenging Cases: What Differentiates Expert From Novice Surgeons? *Journal of Surgical Education*, *78*(2), 450–461. https://doi.org/10.1016/j.jsurg.2020.08.019

Giangreco, A., Carugati, A. and Sebastiano, A. (2010), “Are we doing the right thing? Food for thought on training evaluation and its context”, PersonnelReview, Vol. 39No. 2, pp. 162-177.

Gold, A. H., Malhotra, A., & Segars, A. H. (2001). Knowledge management: An organizational capabilities perspective. *Journal of Management Information Systems*, *18*(1), 185–214. https://doi.org/10.1080/07421222.2001.11045669

Gomez, M. J., & Ruipérez-Valiente, J. A. (2022). Analyzing the Evolution of Digital Assessment in Education Literature Using Bibliometrics and Natural Language Processing. In *Handbook of Research on Digital-Based Assessment and Innovative Practices in Education* (pp. 178–200). IGI Global. https://doi.org/10.4018/978-1-6684-2468-1.ch009

Grandi, F., Peruzzini, M., Cavallaro, S., Prati, E., & Pellicciari, M. (2021). Creation of a UX index to design human tasks and workstations. *International Journal of Computer*

Integrated Manufacturing. https://doi.org/10.1080/0951192X.2021.1972470

Grenier, R. S., & Kehrhahn, M. (2008). Toward an integrated model of expertise redevelopment and its implications for HRD. *Human Resource Development Review*, *7*(2), 198–217. https://doi.org/10.1177/1534484308316653

Guo, H., Li, H., Chan, G., & Skitmore, M. (2012). Using game technologies to improve the safety of construction plant operations. *Accident Analysis & Prevention*, *48*, 204–213. https://doi.org/10.1016/j.aap.2011.06.002

Hair, J. F. J., Page, M., & Brunsveld, N. (2019). Essentials of Business Research Methods (4th editio). Routledge. https://doi.org/10.4324/9780429203374

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (Second ed.). SAGE Publications Inc.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. European Business Review, 31(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hair, J. F., Sarstedt, M., & Ringle, C. M. (2019). Rethinking some of the rethinking of partial least squares. *European Journal of Marketing*, *53*(4), 566–584. https://doi.org/10.1108/EJM-10-2018-0665

Hall-Ellis, S. D., & Grealy, D. S. (2013). *The Dreyfus Model of Skill Acquisition: A Career Development Framework for Succession Planning and Management in Academic Libraries*.

Han, J., & Conti, D. (2020). The use of UTAUT and post acceptance models to investigate the attitude towards a telepresence robot in an educational setting. *Robotics*, *9*(2). https://doi.org/10.3390/ROBOTICS9020034

Han, Y., Diao, Y., Yin, Z., Jin, R., Kangwa, J., & Ebohon, O. J. (2021). Immersive technology-driven investigations on influence factors of cognitive load incurred in construction site hazard recognition, analysis and decision making. *Advanced Engineering Informatics, 48*. https://doi.org/10.1016/j.aei.2021.101298

Hartl, E., & Berger, B. (2017). Escaping reality: Examining the role of presence and escapism in user adoption of virtual reality glasses. *Proceedings of the 25th European Conference on Information Systems, ECIS 2017*, *2017*, 2413–2428.

Havig, P., McIntire, J., & Geiselman, E. (2011). Virtual reality in a cave: limitations and the need for HMDs? In P. L. Marasco & P. R. Havig (Eds.), *Head- and Helmet-Mounted Displays XVI: Design and Applications* (Vol. 8041, p. 804107). SPIE. https://doi.org/10.1117/12.883855

Health and Safety Executive. (1999). *Health and safety in engineering workshops* (2nd ed.).

HSE Books.

Health and Safety Executive. (2014). *Risk assessment A brief guide to controlling risks in the workplace*. www.hse.gov.uk/pubns/indg163.htm

Heilig, M. L. (1962). *Sensorama Simulator U.S. Patent No - 3*.

Heilman, G. E., & Brusa, J. (2006). Validating the End-User Computing Satisfaction Survey Instrument in Mexico. *International Journal of Technology and Human Interaction (IJTHI)*, *2*(4), 84–96. https://doi.org/10.4018/jthi.2006100105

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review*, *33*(3), 405–431. https://doi.org/10.1108/IMR-09-2014-0304

Hernández-Chávez, M., Cortés-Caballero, J. M., Pérez-Martínez, Á. A., Hernández-Quintanar, L. F., Roa-Tort, K., Rivera-Fernández, J. D., & Fabila-Bustos, D. A. (2021). Development of virtual reality automotive lab for training in engineering students. *Sustainability (Switzerland)*, *13*(17). https://doi.org/10.3390/su13179776

Hirsh, W., & Carter, A. (2002). *New directions in management development*. Institute for Employment Studies.

Hjelmervik, K., Nazir, S., & Myhrvold, A. (2018). Simulator training for maritime complex tasks: an experimental study. *WMU Journal of Maritime Affairs*, *17*(1), 17–30. https://doi.org/10.1007/s13437-017-0133-0

Horn, J. L., & Mcardle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3), 117–144. https://doi.org/10.1080/03610739208253916

Hricko, M., & Howell, S. L. (2006). *Online Assessment and Measurement* (M. Hricko & S. L. Howell (Eds.)). IGI Global. https://doi.org/10.4018/978-1-59140-720-1

Hsu, C. L., & Lin, J. C. C. (2008). Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation. *Information and Management*, *45*(1), 65–74. https://doi.org/10.1016/j.im.2007.11.001

Huang, F. H. (2020). Adapting UTAUT2 to assess user acceptance of an e-scooter virtual reality service. *Virtual Reality*, *2001*. https://doi.org/10.1007/s10055-019-00424-7

Hudson, P. (2003). Applying the lessons of high risk industries to health care. *Quality and Safety in Health Care*, *12*(90001), 7i – 12. https://doi.org/10.1136/qhc.12.suppl_1.i7

Hunt, J., & Weintraub, J. (2007). *The Coaching Organization: A Strategy for Developing*

Leaders. SAGE Publications, Inc. https://doi.org/10.4135/9781483329062

Hussein, Z. (2017). Leading to Intention: The Role of Attitude in Relation to Technology Acceptance Model in E-Learning. *Procedia Computer Science*, *105*, 159–164. https://doi.org/10.1016/j.procs.2017.01.196

Imlig-Iten, N., & Petko, D. (2018). Comparing Serious Games and Educational Simulations: Effects on Enjoyment, Deep Thinking, Interest and Cognitive Learning Gains. *Simulation and Gaming*, *49*(4), 401–422. https://doi.org/10.1177/1046878118779088

Isleyen, E., & Duzgun, H. S. (2019). Use of virtual reality in underground roof fall hazard assessment and risk mitigation. *International Journal of Mining Science and Technology*, *29*(4), 603–607. https://doi.org/10.1016/j.ijmst.2019.06.003

Ivaniushin, D. A., Shtennikov, D. G., Efimchick, E. A., & Lyamin, A. V. (2016). Automated Assessment in Massive Open Online Courses. In *International Conferences on Internet Technologies & Society (ITS), Education Technologies (ICEduTECH), and Sustainability, Technology and Education (STE)*.

Jackson, D. (2000). A semi-automated approach to online assessment. *ACM SIGCSE Bulletin*, *32*(3), 164–167. https://doi.org/10.1145/353519.343160

Jain, S., Lee, S., Barber, S. R., Chang, E. H., & Son, Y. J. (2020). Virtual reality based hybrid simulation for functional endoscopic sinus surgery. *IISE Transactions on Healthcare Systems Engineering*, *10*(2), 127–141. https://doi.org/10.1080/24725579.2019.1692263

Jeelani, I., Han, K., & Albert, A. (2020). Development of virtual reality and stereo-panoramic environments for construction safety training. *Engineering, Construction and Architectural Management*, *27*(8), 1853–1876. https://doi.org/10.1108/ECAM-07-2019-0391

Jensen, A. M. D., Jensen, M. M., Korsager, A. S., Ong, M.-S., Magrabi, F., & Coiera, E. (2012). Using Virtual Worlds to Train Healthcare Workers-A Case Study Using Second Life to Improve the Safety of Inpatient Transfers. *Electroonic Journal of Health Informatics*, *7*(1), 7.

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. https://doi.org/10.1007/s10639-017-9676-0

Jin, R., Zou, P. X. W., Piroozfar, P., Wood, H., Yang, Y., Yan, L., & Han, Y. (2019). A science mapping approach based review of construction safety research. *Safety Science*, *113*(September 2018), 285–297. https://doi.org/10.1016/j.ssci.2018.12.006

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*. https://doi.org/10.1016/j.apergo.2020.103286

Kassem, M., Benomran, L., & Teizer, J. (2017). Virtual environments for safety learning in construction and engineering: seeking evidence and identifying gaps for future research. *Visualization in Engineering*, *5*(1). https://doi.org/10.1186/s40327-017-0054-1

Kavanagh, S., Luxton-Reilly, A., Wuensche, B., & Plimmer, B. (2017). A Systematic Review of Virtual Reality in Education. *Themes in Science and Technology Education*, *10*(2), 85–119.

Kazar, G., & Comu, S. (2021). Effectiveness of Serious Games for Safety Training: A Mixed Method Study. *Journal of Construction Engineering and Management*, *147*(8), 04021091. https://doi.org/10.1061/(asce)co.1943-7862.0002119

Kerr, S., Muller, D., McKinon, W., & Mc Inerney, P. (2016). An online formative assessment tool to prepare students for summative assessment in physiology. *African Journal of Health Professions Education*, *8*(1), 72. https://doi.org/10.7196/ajhpe.2016.v8i1.581

Kirkpatrick, D. L. (2006). *Evaluating training programs : the four levels* (3rd ed.). Berrett-Koehler.

Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*.

Kumar, V. V., Carberry, D., Beenfeldt, C., Andersson, M. P., Mansouri, S. S., & Gallucci, F. (2021). Virtual reality in chemical and biochemical engineering education and training. *Education for Chemical Engineers*, *36*, 143–153. https://doi.org/10.1016/j.ece.2021.05.002

Kunz, R. E., & Santomier, J. P. (2019). Sport content and virtual reality technology acceptance. *Sport, Business and Management: An International Journal*. https://doi.org/10.1108/SBM-11-2018-0095

Kupfer, A., Ableitner, L., Schöb, S., & Tiefenbeck, V. (2016). Decision criteria for technology usage behavior Technology Adoption vs. Continuous Usage Intention: do Decision Criteria Change when Using a Technology? *Americas Conference on Information Systems*, *August*.

Kwegyir-Afful, E., Hassan, T. O., & Kantola, J. I. (2021). Simulation-based assessments of fire emergency preparedness and response in virtual reality. *International Journal of Occupational Safety and Ergonomics*. https://doi.org/10.1080/10803548.2021.1891395

Lai, C., Chen, X., Chen, X., Wang, Z., Wu, X., & Zhao, S. (2015). A fuzzy comprehensive evaluation model for flood risk based on the combination weight of game theory. *Natural Hazards*, *77*(2), 1243–1259. https://doi.org/10.1007/s11069-015-1645-6

Lasunon, P. (2019). Triangular fuzzy grading method and its software. *2019 5th International*

Conference on Engineering, Applied Sciences and Technology (ICEAST), 1–4. https://doi.org/10.1109/ICEAST.2019.8802546

Le, Q. T., Pedro, A., & Park, C. S. (2015). A Social Virtual Reality Based Construction Safety Education System for Experiential Learning. *Journal of Intelligent & Robotic Systems*, *79*(3–4), 487–506. https://doi.org/10.1007/s10846-014-0112-z

Leder, J., Horlitz, T., Puschmann, P., Wittstock, V., & Schütz, A. (2019). Comparing immersive virtual reality and powerpoint as methods for delivering safety training: Impacts on risk perception, learning, and decision making. *Safety Science*, *111*(July), 271–286. https://doi.org/10.1016/j.ssci.2018.07.021

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255–275. https://doi.org/10.1037/0033-2909.125.2.255

Lew, S., Tan, G. W. H., Loh, X. M., Hew, J. J., & Ooi, K. B. (2020). The disruptive mobile wallet in the hospitality industry: An extended mobile technology acceptance model. *Technology in Society*, *63*. https://doi.org/10.1016/j.techsoc.2020.101430

Li, H., Chan, G., & Skitmore, M. (2012). Multiuser Virtual Safety Training System for Tower Crane Dismantlement. *Journal of Computing in Civil Engineering*, *26*(5), 638–647. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000170

Li, J. P., & Kishore, R. (2006). How Robust is the UTAUT Instrument? A Multigroup Invariance Analysis in the Context of Acceptance and Use of Online Community Weblog Systems. *SIGMIS CPR '06 Proceedings of the 2006 ACM SIGMIS CPR Conference on Computer Personnel Research: Forty Four Years of Computer Personnel Research: Achievements, Challenges & the Future*, 183–189. https://doi.org/10.1145/1125170.1125218

Li, M. kun, Liu, Y. kuo, Peng, M. jun, Xie, C. li, & Yang, L. qun. (2018). The digital simulation and fuzzy evaluation to reduce the likelihood of unsafe behavior in nuclear decommissioning. *Annals of Nuclear Energy*, *119*, 331–341. https://doi.org/10.1016/j.anucene.2018.05.023

Li, S., Cui, J., Hao, A., Zhang, S., & Zhao, Q. (2021). Design and Evaluation of Personalized Percutaneous Coronary Intervention Surgery Simulation System. *IEEE Transactions on Visualization and Computer Graphics*, *27*(11), 4150–4160. https://doi.org/10.1109/TVCG.2021.3106478

Liang, Z., Zhou, K., & Gao, K. (2019). Development of Virtual Reality Serious Game for Underground Rock-Related Hazards Safety Training. *IEEE Access*, *7*, 118639–118649. https://doi.org/10.1109/ACCESS.2019.2934990

Lin, K.-Y., Lee, W., Azari, R., & Migliaccio, G. C. (2018). Training of Low-Literacy and

Low-English-Proficiency Hispanic Workers on Construction Fall Fatality. *Journal of Management in Engineering*, *34*(2), 05017009. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000573

Lin, K.-Y., Son, J., & Rojas, E. M. (2011). A pilot study of a 3D game environment for construction safety education. *Journal of Information Technology in Construction (ITcon)*, *16*, 69–84.

Liu, Y., Eckert, C. M., & Earl, C. (2020). A review of fuzzy AHP methods for decision-making with subjective judgements. In *Expert Systems with Applications* (Vol. 161). Elsevier Ltd. https://doi.org/10.1016/j.eswa.2020.113738

Loh, C. S. (2011). Using in situ data collection to improve the impact and return of investment of game-based learning. *Proceedings of the 61st International Council for Educational Media and the XIII International Symposium on Computers in Education (ICEM&SIIE'2011) Joint Conference*, *May*, 801–811.

Loh, C. S. (2012). Information Trails: In-Process Assessment of Game-Based Learning. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in Game-Based Learning: Foundations, Innovations, and Perspectives* (pp. 123–144). Springer New York. https://doi.org/10.1007/978-1-4614-3546-4_8

Loh, C. S. (2013). Improving the impact and return of investment of game-based learning. *International Journal of Virtual and Personal Learning Environments*, *4*(1), 1–15. https://doi.org/10.4018/jvple.2013010101

Loh, C. S., & Sheng, Y. (2013). Performance metrics for serious games: Will the (real) expert please step forward? *Proceedings of CGAMES 2013 USA - 18th International Conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational and Serious Games*, *July*, 202–206. https://doi.org/10.1109/CGames.2013.6632633

Loh, C. S., Sheng, Y., & Li, I. H. (2015). Predicting expert-novice performance as serious games analytics with objective-oriented and navigational action sequences. *Computers in Human Behavior*, *49*, 147–155. https://doi.org/10.1016/j.chb.2015.02.053

Lyon, L. J. (2014). Development of teaching expertise viewed through the Dreyfus Model of Skill Acquisition. *Journal of the Scholarship of Teaching and Learning*, 88–105. https://doi.org/10.14434/josotl.v15i1.12866

Ma, L., Liu, Y., & Zhou, X. P. (2010). Fuzzy comprehensive evaluation method of F statistics weighting in identifying mine water inrush source. In *International Journal of Engineering, Science and Technology* (Vol. 2, Issue 7).

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, *35*(6), 691–707. https://doi.org/10.1111/jcal.12375

Manca, D., Brambilla, S., & Colombo, S. (2013). Bridging between Virtual Reality and accident simulation for training of process-industry operators. *Advances in Engineering Software*, *55*, 1–9. https://doi.org/10.1016/j.advengsoft.2012.09.002

Mao, E., & Palvia, P. (2008). Exploring the effects of direct experience on IT use: An organizational field study. *Information and Management*, *45*(4), 249–256. https://doi.org/10.1016/j.im.2008.02.007

Marakarkandy, B., & Yajnik, N. (2013). Re-examining and empirically validating the End User Computing Satisfaction models for satisfaction measurement in the internet banking context. *International Journal of Bank Marketing*, *31*(6), 440–455. https://doi.org/10.1108/IJBM-06-2013-0051

Marques, B. P., Villate, J. E., & Carvalho, C. V. (2011). Applying the UTAUT model in engineering higher education: Teacher's technology adoption. *Proceedings of the 6th Iberian Conference on Information Systems and Technologies, CISTI 2011*, 1–6.

Matthews, L. (2017). Applying Multigroup Analysis in PLS-SEM: A Step-by-Step Process. In H. Latan & R. Noonan (Eds.), *Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications* (pp. 219–243). Springer International Publishing. https://doi.org/10.1007/978-3-319-64069-3_10

Menin, A., Torchelsen, R., & Nedel, L. (2018). An Analysis of VR Technology Used in Immersive Simulations with a Serious Game Perspective. *IEEE Computer Graphics and Applications*, *38*(2), 57–73. https://doi.org/10.1109/MCG.2018.021951633

Merhi, M., Hone, K., & Tarhini, A. (2019). A cross-cultural study of the intention to use mobile banking between Lebanese and British consumers: Extending UTAUT2 with security, privacy and trust. *Technology in Society*, *59*(June), 101151. https://doi.org/10.1016/j.techsoc.2019.101151

Min, W., Frankosky, M., Mott, B. W., Rowe, J., Smith, P. A. M., Wiebe, E., Boyer, K., & Lester, J. (2019). DeepStealth: Game-Based Learning Stealth Assessment with Deep Neural Networks. *IEEE Transactions on Learning Technologies*, *1382*(c). https://doi.org/10.1109/TLT.2019.2922356

Mislevy, R. J., & Haertel, G. D. (2007). Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice*, *25*(4), 6–20. https://doi.org/10.1111/j.1745-3992.2006.00075.x

Mislevy, R. J., Haertel, G., Riconscente, M., Rutstein, D. W., & Ziker, C. (2017). Evidence-Centered Assessment Design. In *Assessing Model-Based Reasoning using Evidence-*

Centered Design (pp. 19–24). SpringerBriefs in Statistics. Springer, Cham. https://doi.org/10.1007/978-3-319-52246-3_3

Mislevy, R. J., & Riconscente, M. (2005). *Evidence-Centered Assessment Design: Layers, Structures, and Terminology* (Issue PADI Technical Report 9).

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). Focus Article: On the Structure of Educational Assessments. *Measurement: Interdisciplinary Research & Perspective*, *1*(1), 3–62. https://doi.org/10.1207/S15366359MEA0101_02

Molnár, G. (2017). How digital and virtual life trapped in visuality and new media systems affect teaching and the learning process. *Virtual Reality - Real Visuality: Virtual, Visual, Veridical, November 2016*, 101–118. https://doi.org/10.3726/b11592

Mondragón-Bernal, I. F. (2020). A system for immersive medical and engineering training based on serious games. *Ingenieria y Universidad*, *24*. https://doi.org/10.11144/JAVERIANA.IUED24.SIME

Muhanna, M. A. (2015). Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions. *Journal of King Saud University - Computer and Information Sciences*, *27*(3), 344–361. https://doi.org/10.1016/j.jksuci.2014.03.023

Nawi, F. A. M., Tambi, A. M. A., Samat, M. F., & Mustapha, W. M. W. (2020). A review on the internal consistency of a scale: the empirical example of the influence of human capital investment on malcom baldridge quality principles in TVET institutions. *Asian People Journal (APJ)*, *3*(1), 19–29. https://doi.org/10.37231/apj.2020.3.1.121

Nazir, S., Sorensen, L. J., Øvergård, K. I., & Manca, D. (2015). Impact of training methods on Distributed Situation Awareness of industrial operators. *Safety Science*, *73*, 136–145. https://doi.org/10.1016/j.ssci.2014.11.015

Nazir, S., Totaro, R., Brambilla, S., Colombo, S., & Manca, D. (2012). Virtual Reality and Augmented-Virtual Reality as Tools to Train Industrial Operators. In *Computer Aided Chemical Engineering* (Vol. 30, Issue December, pp. 1397–1401). https://doi.org/10.1016/B978-0-444-59520-1.50138-X

Neitzel, D. K. (2018). Identifying the Requirements for Qualified, Unqualified, and Competent Persons Electrical Safety Training. IEEE Transactions on Industry Applications, 54(1), 5–9. https://doi.org/10.1109/TIA.2017.2741445

Nichols, J. T. (2009). The 3 Directions: Situated Information Literacy.

Nicoletti, L., & Padovano, A. (2019). Human factors in occupational health and safety 4.0: a cross-sectional correlation study of workload, stress and outcomes of an industrial emergency response. *International Journal of Simulation and Process Modelling*, *14*(2), 178. https://doi.org/10.1504/IJSPM.2019.099912

Numfu, M., Riel, A., & Noel, F. (2020). Virtual Reality Technology for Maintenance Training. *Applied Science and Engineering Progress*, *13*(3), 1–9. https://doi.org/10.14416/j.asep.2020.04.008

Nykänen, M., Puro, V., Tiikkaja, M., Kannisto, H., Lantto, E., Simpura, F., Uusitalo, J., Lukander, K., Räsänen, T., Heikkilä, T., & Teperi, A.-M. (2020). Implementing and evaluating novel safety training methods for construction sector workers: Results of a randomized controlled trial. *Journal of Safety Research*, *75*, 205–221. https://doi.org/10.1016/j.jsr.2020.09.015

O.Nyumba, T., Wilson, K., Derrick, C. J., & Mukherjee, N. (2018). The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution*, *9*(1), 20–32. https://doi.org/10.1111/2041-210X.12860

Odu, G. O. (2019). Weighting methods for multi-criteria decision making technique. *Journal of Applied Sciences and Environmental Management*, *23*(8), 1449. https://doi.org/10.4314/jasem.v23i8.7

Ogbuanya, T. C., & Onele, N. O. (2018). Investigating the Effectiveness of Desktop Virtual Reality for Teaching and Learning of Electrical/Electronics Technology in Universities. *Computers in the Schools*, *35*(3), 226–248. https://doi.org/10.1080/07380569.2018.1492283

Ogourtsova, T., Archambault, P. S., & Lamontagne, A. (2019). Exploring barriers and facilitators to the clinical use of virtual reality for post-stroke unilateral spatial neglect assessment. *Disability and Rehabilitation*, *41*(3), 284–292. https://doi.org/10.1080/09638288.2017.1387292

Ojados Gonzalez, D., Martin-Gorriz, B., Ibarra Berrocal, I., Macian Morales, A., Adolfo Salcedo, G., & Miguel Hernandez, B. (2017). Development and assessment of a tractor driving simulator with immersive virtual reality for training to avoid occupational hazards. *Computers and Electronics in Agriculture*, *143*, 111–118. https://doi.org/10.1016/j.compag.2017.10.008

Oliveri, M. E., Lawless, R., & Mislevy, R. J. (2019). Using Evidence-Centered Design to Support the Development of Culturally and Linguistically Sensitive Collaborative Problem-Solving Assessments. International Journal of Testing, 19(3), 270–300. https://doi.org/10.1080/15305058.2018.1543308

Osti, F., de Amicis, R., Sanchez, C. A., Tilt, A. B., Prather, E., & Liverani, A. (2021). A VR training system for learning and skills development for construction workers. *Virtual Reality*, *25*(2), 523–538. https://doi.org/10.1007/s10055-020-00470-6

Patle, D. S., Ahmad, Z., & Rangaiah, G. P. (2014). Operator training simulators in the chemical industry: review, issues, and future directions. *Reviews in Chemical Engineering*, *30*(2), 199–216. https://doi.org/10.1515/revce-2013-0027

Patle, D. S., Manca, D., Nazir, S., & Sharma, S. (2019). Operator training simulators in virtual reality environment for process operators: a review. *Virtual Reality*, *23*(3), 293–311. https://doi.org/10.1007/s10055-018-0354-3

Pedro, A., Pham, H. C., Kim, J. U., & Park, C. (2020). Development and evaluation of context-based assessment system for visualization-enhanced construction safety education. *International Journal of Occupational Safety and Ergonomics*, *26*(4), 811–823. https://doi.org/10.1080/10803548.2018.1553377

Peña, A. (2010). The Dreyfus model of clinical problem-solving skills acquisition: a critical perspective. *Medical Education Online*, *15*, 1–11. https://doi.org/10.3402/meo.v15i0.4846

Perlman, A., Sacks, R., & Barak, R. (2014). Hazard recognition and risk perception in construction. *Safety Science*, *64*, 13–21. https://doi.org/10.1016/j.ssci.2013.11.019

Persky, A. M., & Robinson, J. D. (2017). Moving from novice to expertise and its implications for instruction. In *American Journal of Pharmaceutical Education* (Vol. 81, Issue 9, pp. 72–80). American Association of Colleges of Pharmacy. https://doi.org/10.5688/ajpe6065

Pham, H. C., Dao, N. N., Cho, S., Nguyen, P. T., & Pham-Hang, A. T. (2019). Construction hazard investigation leveraging object anatomization on an augmented photoreality platform. *Applied Sciences (Switzerland)*, *9*(21). https://doi.org/10.3390/app9214477

Polivka, B. J., Anderson, S., Lavender, S. A., Sommerich, C. M., Stredney, D. L., Wills, C. E., & Darragh, A. R. (2019). Efficacy and Usability of a Virtual Simulation Training System for Health and Safety Hazards Encountered by Healthcare Workers. *Games for Health Journal*, *8*(2), 121–128. https://doi.org/10.1089/g4h.2018.0068

Poyade, M., Eaglesham, C., Trench, J., & Reid, M. (2021). A Transferable Psychological Evaluation of Virtual Reality Applied to Safety Training in Chemical Manufacturing. *Journal of Chemical Health and Safety*, *28*(1), 55–65. https://doi.org/10.1021/acs.chas.0c00105

Puteh, F., Bin, M. H., & Ong, A. (2017). Quantitative Data Analysis:Choosing Between SPSS, PLS and AMOS in Social Science Research. *International Interdisciplinary Journal of Scientific Research*.

Rafiq, M. (2015). Training Evaluation in an Organization using Kirkpatrick Model: A Case Study of PIA. *Journal of Entrepreneurship & Organization Management*, *04*(03).

https://doi.org/10.4172/2169-026x.1000151

Ragan, E. D., Sowndararajan, A., Kopper, R., & Bowman, D. (2010). The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments. Presence: Teleoperators and Virtual Environments, 19(6), 527–543. https://doi.org/10.1162/pres_a_00016

Rahouti, A., Lovreglio, R., Datoussaïd, S., & Descamps, T. (2021). Prototyping and Validating a Non-immersive Virtual Reality Serious Game for Healthcare Fire Safety Training. *Fire Technology*, *57*(6), 3041–3078. https://doi.org/10.1007/s10694-021-01098-x

Ramayah, T., Ignatius, J., & Aafaqi, B. (2005). PC Usage among students in a private institution of higher learning: The moderating role of prior experience. *Malaysian Journal of Educators and Education*, *20*, 131–152.

Ramírez-Correa, P., Arenas-Gaitán, J., & Rondán-Cataluña, F. J. (2015). Gender and acceptance of e-learning: A multi-group analysis based on a structural equation model among college students in Chile and Spain. *PLoS ONE*, *10*(10), 1–17. https://doi.org/10.1371/journal.pone.0140460

Ringle, C. M., Wende, S., & Becker, J.-M. (2015). *SmartPLS 3*. SmartPLS GmbH, Boenningstedt. http://www.smartpls.com

Riva, G., Wiederhold, B. K., & Gaggioli, A. (2016). Being different: The transformative potential of virtual reality. *Annual Review of CyberTherapy and Telemedicine*, *14*, 3–6.

Rondan-Cataluña, F. J., Arenas-Gaitán, J., & Ramírez-Correa, P. E. (2015). A comparison of the different versions of popular technology acceptance models a non-linear perspective. *Kybernetes*, *44*(5), 788–805. https://doi.org/10.1108/K-09-2014-0184

Sacks, R., Perlman, A., & Barak, R. (2013). Construction safety training using immersive virtual reality. *Construction Management and Economics*, *31*(9), 1005–1017. https://doi.org/10.1080/01446193.2013.828844

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. In *Instructional Science* (Vol. 18).

SAGE Research Methods Datasets. (2019). Learn About Structural Equation Modeling in SmartPLS With Data From the Customer Behavior in Electronic Commerce Study in Ecuador (2017).

Sakowitz, S. M., Inglehart, M. R., Ramaswamy, V., Edwards, S., Shoukri, B., Sachs, S., & Kim-Berman, H. (2020). A comparison of two-dimensional prediction tracing and a virtual reality patient methods for diagnosis and treatment planning of orthognathic cases in dental students: a randomized preliminary study. *Virtual Reality*, *24*(3), 399–409.

https://doi.org/10.1007/s10055-019-00413-w

Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The Science of Training and Development in Organizations: What Matters in Practice. *Psychological Science in the Public Interest, Supplement*, *13*(2), 74–101. https://doi.org/10.1177/1529100612436661

Serpa, Y. R., Nogueira, M. B., Rocha, H., Macedo, D. V., & Rodrigues, M. A. F. (2020). An interactive simulation-based game of a manufacturing process in heavy industry. *Entertainment Computing*, *34*, 100343. https://doi.org/10.1016/j.entcom.2020.100343

Shamsuzzoha, A., Toshev, R., Vu Tuan, V., Kankaanpaa, T., & Helo, P. (2019). Digital factory–virtual reality environments for industrial training and maintenance. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2019.1628072

Sharma, A. K., & Kumar, D. (2012). User acceptance of desktop based computer software using UTAUT model and addition of new moderators. *International Journal Of Computer Science & Engineering Technology*, *3*(10), 509–515.

Shen, C., Ho, J., Ly, P. T. M., & Kuo, T. (2019). Behavioural intentions of using virtual reality in learning: perspectives of acceptance of information technology and learning style. *Virtual Reality*, *23*(3), 313–324. https://doi.org/10.1007/s10055-018-0348-1

Shi, Y., Du, J., Ahn, C. R., & Ragan, E. (2019). Impact assessment of reinforced learning methods on construction workers' fall risk behavior using virtual reality. *Automation in Construction*, *104*(October 2018), 197–214. https://doi.org/10.1016/j.autcon.2019.04.015

Shi, Y., Du, J., & Worthy, D. A. (2020). The impact of engineering information formats on learning and execution of construction operations: A virtual reality pipe maintenance experiment. *Automation in Construction*, *119*, 103367. https://doi.org/10.1016/j.autcon.2020.103367

Shute, V. J., & Emihovich, B. (2018). Assessing Problem-Solving Skills in Game-Based Immersive Environments. In J. Voogt (Ed.), *Second Handbook of Information Technology in Primary and Secondary Education* (pp. 635–648). Springer International Publishing. https://doi.org/10.1007/978-3-319-71054-9_40

Shute, V. J., & Rahimi, S. (2021). Stealth assessment of creativity in a physics video game. *Computers in Human Behavior*, *116*. https://doi.org/10.1016/j.chb.2020.106647

Shute, V. J., & Wang, L. (2016). Assessing and Supporting Hard-to-Measure Constructs in Video Games. In *The Handbook of Cognition and Assessment* (pp. 535–562). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118956588.ch22

Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, *63*, 106–117. https://doi.org/10.1016/j.chb.2016.05.047

Sitar-Tăut D. A. (2021). Mobile learning acceptance in social distancing during the COVID-19 outbreak: The mediation effect of hedonic motivation. Human behavior and emerging technologies, 3(3), 366–378. https://doi.org/10.1002/hbe2.261

Snow, E., Haertel, G., Fulkerson, D., Mingyu Feng, P., & Nichols, P. (2010). *Leveraging Evidence-Centered Assessment Design in Large-Scale and Formative Assessment Practices*.

Somrak, A., Humar, I., Hossain, M. S., Alhamid, M. F., Hossain, M. A., & Guna, J. (2019). Estimating VR Sickness and user experience using different HMD technologies: An evaluation study. Future Generation Computer Systems, 94, 302–316. https://doi.org/10.1016/j.future.2018.11.041

Song, H., Kim, T., Kim, J., Ahn, D., & Kang, Y. (2021). Effectiveness of VR crane training with head-mounted display: Double mediation of presence and perceived usefulness. *Automation in Construction*, *122*. https://doi.org/10.1016/j.autcon.2020.103506

Soret, R., Charras, P., Hurter, C., & Peysakhovich, V. (2019). *Attentional orienting in virtual reality using endogenous and exogenous cues in auditory and visual modalities*. 1–8. https://doi.org/10.1145/3314111.3321490

Souza, D. M., Felizardo, K. R., & Barbosa, E. F. (2016). A systematic literature review of assessment tools for programming assignments. *Proceedings - 2016 IEEE 29th Conference on Software Engineering Education and Training, CSEEandT 2016*, 147–156. https://doi.org/10.1109/CSEET.2016.48

Stransky, J., Bassett, L., Bodnar, C., Anastasio, D., Burkey, D., & Cooper, M. (2021). A retrospective analysis on the impacts of an immersive digital environment on chemical engineering students' moral reasoning. *Education for Chemical Engineers*, *35*, 22–28. https://doi.org/10.1016/j.ece.2020.12.003

Sudhagar, C., & Ganesan, K. (2011). *A Fuzzy Approach to Educational Grading System*.

Sutherland, I. E. (1965). The Ultimate Display. *Proceedings of IFIP Congress*, 506–508.

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Research in Science Education, 48(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tawadrous, M., Rojas, D., Kapralos, B., Hogue, A., & Dubrowski, A. (2017). The effects of stereoscopic 3D on knowledge retention within a serious gaming environment. *Multimedia Tools and Applications*, *76*(5), 7301–7319. https://doi.org/10.1007/s11042-016-3394-2

Tehreem, Y., Fracaro, S. G., Gallagher, T., Toyoda, R., Bernaerts, K., Glassey, J., Abegao, F.

R., Wachsmuth, S., Wilk, M., & Pfeiffer, T. (2022). May I Remain Seated: A Pilot Study on the Impact of Reducing Room-Scale Trainings to Seated Conditions for Long Procedural Virtual Reality Trainings. In E. Wu, Z. Pan, & A. Song (Eds.), *2022 8th International Conference on Virtual Reality (ICVR)* (pp. 62–71). IEEE. https://doi.org/10.1109/ICVR55215.2022.9848222

Teo, H.-H., Oh, L.-B., Liu, C., & Wei, K.-K. (2003). An empirical study of the effects of interactivity on web user attitude. *International Journal of Human-Computer Studies*, *58*(3), 281–305. https://doi.org/10.1016/S1071-5819(03)00008-9

Teo, T., & Noyes, J. (2014). Explaining the intention to use technology among pre-service teachers: a multi-group analysis of the Unified Theory of Acceptance and Use of Technology. *Interactive Learning Environments*, *22*(1), 51–66. https://doi.org/10.1080/10494820.2011.641674

Thompson, M., & Falco, J. (2020). Leveraging technology to minimize operational error and hazards—Rise of the Operator 2.0. *Process Safety Progress*, *39*(1), 1–5. https://doi.org/10.1002/prs.12078

Tian, Y., Liu, H., Yin, J., Luo, M., & Wu, G. (2015). Evaluation of simulation-based training for aircraft carrier marshalling with learning cubic and Kirkpatrick's models. *Chinese Journal of Aeronautics*, *28*(1), 152–163. https://doi.org/10.1016/j.cja.2014.12.002

Toyoda, R., Russo Abegão, F., Gill, S., & Glassey, J. (2021). Drivers of immersive virtual reality adoption intention: a multi-group analysis in chemical industry settings. *Virtual Reality*. https://doi.org/10.1007/s10055-021-00586-3

Tsai, W. S., Yang, Q., & Liu, Y. (2013). Young Chinese Consumers' Snob and Bandwagon Luxury Consumption Preferences. *Journal of International Consumer Marketing*, *25*(5), 290–304. https://doi.org/10.1080/08961530.2013.827081

Tynjälä, P. (2008). Perspectives into learning at the workplace. In *Educational Research Review* (Vol. 3, Issue 2, pp. 130–154). https://doi.org/10.1016/j.edurev.2007.12.001

Udeozor, C., Toyoda, R., Russo Abegão, F., & Glassey, J. (2021). Perceptions of the use of virtual reality games for chemical engineering education and professional training. *Higher Education Pedagogies*, *6*(1), 175–194. https://doi.org/10.1080/23752696.2021.1951615

Unsworth, C. A. (2001). The clinical reasoning of novice and expert occupational therapists. *Scandinavian Journal of Occupational Therapy*, *8*(4), 163–173. https://doi.org/10.1080/110381201317166522

Vahdatikhaki, F., El Ammari, K., Langroodi, A. K., Miller, S., Hammad, A., & Doree, A. (2019). Beyond data visualization: A context-realistic construction equipment training simulators. *Automation in Construction*, *106*. https://doi.org/10.1016/j.autcon.2019.102853

Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis. *Review of Educational Research*, *85*(4), 475–511. https://doi.org/10.3102/0034654314564881

Van Slyke, C., Ilie, V., Lou, H., & Stafford, T. (2007). Perceived critical mass and the adoption of a communication technology. *European Journal of Information Systems*, *16*(3), 270–283. https://doi.org/10.1057/palgrave.ejis.3000680

van Wyk, E. A., & de Villiers, M. R. (2019). An evaluation framework for virtual reality safety training systems in the South African mining industry. *Journal of the Southern African Institute of Mining and Metallurgy*, *119*(5), 427–436. https://doi.org/10.17159/2411-9717/53/2019

Vaquero-álvarez, E., Cubero-Atienza, A., Martínez-Jiménez, M. P., Vaquero-Abellán, M., Redel-Macías, M. D., & Aparicio-Martínez, P. (2020). Occupational safety and health training for undergraduates nursing students: A spanish pilot. *International Journal of Environmental Research and Public Health*, *17*(22), 1–14. https://doi.org/10.3390/ijerph17228381

Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology*, *18*(1). https://doi.org/10.1186/s12874-018-0594-7

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, *39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, *36*(1), 157–178. https://doi.org/10.2307/41410412

Venkatesh, V., Thong, J., & Xu, X. (2016). Unified Theory of Acceptance and Use of

Technology: A Synthesis and the Road Ahead. *Journal of the Association for Information Systems*, *17*(5), 328–376. https://doi.org/10.17705/1jais.00428

Viitaharju, P., Yliniemi, K., Nieminen, M., & Karttunen, A. J. (2021). Learning experiences from digital laboratory safety training. *Education for Chemical Engineers*, *34*, 87–93. https://doi.org/10.1016/j.ece.2020.11.009

Voorhis, V. Van, & Paris, B. (2019). Simulations and Serious Games: Higher Order Thinking Skills Assessment. *Journal of Applied Testing Technology*, *20*(S1), 35–42.

Wallace, C. A., Pontin, D., Dokova, K., Mikkonen, I., Savage, E., & Koskinen, L. (2018). Developing and Translating a New Model for Teaching Empowerment Into Routine Chronic Care Management. *Journal of Patient Experience*, *5*(1), 34–42. https://doi.org/10.1177/2374373517721516

Wang, P., Wu, P., Chi, H.-L., & Li, X. (2020). Adopting lean thinking in virtual reality-based personalized operation training using value stream mapping. *Automation in Construction*, *119*, 103355. https://doi.org/10.1016/j.autcon.2020.103355

Xiao, L., & Dasgupta, S. (2002). Measurement of user satisfaction with web-based information systems: an empirical study. *Americas Conference on Information Systems*, 1149–1155.

Xu, Z., & Zheng, N. (2021). Incorporating Virtual Reality Technology in Safety Training Solution for Construction Site of Urban Cities. *Sustainability*, *13*(1), 243. https://doi.org/10.3390/su13010243

Yılmaz, F. G. K., Ustun, A. B., & Yılmaz, R. (2020). Investigation of Pre-Service Teachers' Opinions on Advantages and Disadvantages of Online Formative Assessment: An Example of Online Multiple-Choice Exam. *Journal of Teacher Education and Lifelong Learning*, *2*(1), 10–19.

Yusof, N., Zin, N. A. M., & Adnan, N. S. (2012). Java Programming Assessment Tool for Assignment Module in Moodle E-learning System. *Procedia - Social and Behavioral Sciences*, *56*, 767–773. https://doi.org/10.1016/j.sbspro.2012.09.714

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zainab, A. M., Kiran, K., Karim, N. H. A., & Sukmawati, M. (2018). UTAUT'S performance consistency: Empirical evidence from a library management system. *Malaysian Journal of Library and Information Science*, *23*(1), 17–32. https://doi.org/10.22452/mjlis.vol23no1.2

Zare, M., Larique, M., Chevriau, S., & Sagot, J. C. (2019). Application of virtual reality to improve physical ergonomics in a control room of a chemical industry. *Advances in*

Intelligent Systems and Computing, *822*, 296–308. https://doi.org/10.1007/978-3-319-96077-7_31

Zhou, M., Long, P., Kong, N., Zhao, L., Jia, F., & Campy, K. S. (2021). Characterizing the motivational mechanism behind taxi driver's adoption of electric vehicles for living: Insights from China. Transportation Research Part A: Policy and Practice, 144, 134–152. https://doi.org/10.1016/j.tra.2021.01.001

Zieky, M. J. (2014). An introduction to the use of evidence-centered design in test development. *Psicología Educativa*, *20*(2), 79–87. https://doi.org/10.1016/j.pse.2014.11.003

Appendix A – Online questionnaire used in the UTAUT2 study (English)

Health and Safety (H&S) Training using Virtual Reality (VR) Game

What is the purpose of this survey?

The aim of this project is to evaluate the acceptance intention as well as continuous usage intention of technology (Virtual reality game) in health and safety training.

Who is conducting this research?

This research is being conducted by Ryo Toyoda under the supervision of Prof. Jarka Glassey, Dr. Fernando Russo Abegão, and Dr. Sue Gill of Newcastle University as part of the European Training Network for Chemical Engineering Immersive Learning (ETN-CHARMING) project.

How specifically do I participate?

You are invited to take part in an online survey. The survey should take less than 15 minutes of your time. The questions will concern your perception on using the virtual reality (VR) game for employee H&S training in chemical industry.

What will you do with my answers to the questionnaire?

Please be assured that the responses you provide are completely anonymous and confidential. Only people directly involved in this survey will have access to the raw data we collect. There is no risk in taking part in this survey. The information will be stored securely on the protected file store service. The survey outcome and report will not include reference to any individual. The anonymized data and resulting analysis may be used in academic and non-academic publications and presentation.

What if I change my mind or have questions?

You can ask questions, stop your participation, or withdraw completely from the research project at any time.

After reading this information, and if you decide to take part in the project, you will be asked to complete the attached informed consent form at the beginning of the survey.

If you have any questions, please don't hesitate to contact:

Ryo Toyoda
ryo.toyoda@newcastle.ac.uk
PhD | School of Engineering | Newcastle University
Merz Court | Newcastle upon Tyne | NE1 7RU | UK

Employee Consent form for a research survey as part of the project

Health and Safety (H&S) Training using Virtual Reality (VR) Game

I confirm that:

- I have read and understood the information about the project, as provided in the Information Sheet.
- I have been given the opportunity to ask questions about the project and my participation in it.
- I voluntarily agree to my participation.
- I understand that I can withdraw my participation to the survey at any time without giving a reason and will not be questioned on why I have withdrawn.
- I understand that other researchers from the Newcastle University and the CHARMING consortium who are part of the project team may have access to the anonymized information collected and that the researchers will follow the agreed procedures for the storage of the data as detailed on the Information Sheet.
- I understand that I will be answering an online survey for the study and my answers will be stored securely on the protected file store service.
- I understand that anonymized answers from the survey may be used in written academic and non-academic publications and conference presentations about the research project.

☐ I have read and agree to the terms and conditions. (Please tick box)

Section A: Demographic profile

Company: ____________

Job Description: ____________

Gender
- Female
- Male

What is your age group?
- Under 20
- 20-29
- 30-39
- 40-49
- 50-59
- 60 & above

Country of Origin: _____________

What is your highest educational qualification or equivalent?
- Primary School
- Middle High School
- Secondary School
- Bachelor's Degree
- Master's Degree
- Doctoral Degree
- Others: ______________

How long have you been working with the company?
- Less than a year
- 1-5 years
- 6-20 years
- More than 20 years

What type of H&S training is available in your company? (You can choose more than one option)
- Formal Classroom setting (Lectures)
- Informal Classroom setting (*e.g.* small group discussion, case studies)
- Coaching or on-the-job training
- Open and distance learning/education (*e.g.* Video Presentation, online platform)
- Virtual Reality (VR) and/or Augmented Reality (AR) environment setting
- Others _______________

Section B: Virtual Reality (VR) / Game Experience

Have you tried using Virtual Reality head-mounted display before?
- Yes
- No

Which VR gaming console(s) or equivalent have you used? (You can choose more than one option)

- Sony PlayStation VR
- HTC Vive
- Oculus Rift/quest
- Samsung Gear VR
- Google Cardboard
- LG 360 VR
- Google Daydream
- None
- Others: ________

How many times have you used it in a year?

- Never
- Once
- 2-5 times
- Over 5 times

Do you play video games?

- Yes
- No

Which gaming device(s) or equivalent have you used? (You can choose more than one option)

- Nintendo (Gameboy, 3DS, WII, Switch, *etc.*)
- PlayStation
- Xbox
- Mobile Phone Games
- Computer Games
- None
- Others: ________

How often do you play per week?

- Never
- Less than 5 hours
- 5-15 hours
- 16-25 hours
- More than 25 hours

Section C: This section tries to understand your perception of using the virtual reality (VR) game for employee H&S training in chemical industry

Items	6-point Likert Scale					
	Strongly Agree	Agree	Slightly Agree	Slightly Disagree	Disagree	Strongly Disagree
1. I think that using the VR environment will be useful for practicing H&S procedures.						

2. Using VR environment will probably enable me to learn the H&S procedures more quickly.						
3. If I use this VR environment, I will improve my performance on H&S procedures.						
4. I think using the VR environment will be clear and understandable.						
5. I think that it will be easy for me to operate the platform in which the VR environment is running.						
6. I think that it will take too long to learn how to use the VR environment to make it worth the effort.						
7. I think that the organization will support me in learning how to use the VR environment.						
8. People who influence my behavior at work think that I should use this VR environment.						
9. I think my supervisor will be very supportive of the use of this VR environment for my job.						
10. I feel that it will be a bad idea to use the VR environment for H&S training.						
11. I think that the actual process of using the VR environment for H&S training is fun.						
12. I think that using VR environment for H&S training will be very frustrating.						
13. If made available to me, I would recommend						

using the VR environment for learning to apply the H&S procedures to my colleagues.						
14. If made available to me, I plan to continue to use VR environment for H&S training frequently.						
15. I think that after using the VR for H&S training, I will be ready to use this learning environment for another training course.						

Appendix B – Online questionnaire used in the UTAUT2 study (German)

Gesundheits- und Sicherheitstraining (G&S) mit Virtual Reality (VR) Spiel

Was ist das Ziel dieser Umfrage?

Das Ziel dieses Projekts ist es, die Akzeptanz und kontinuierliche Verwendung von Technologie (Virtual Reality Spiel) im Gesundheits- und Sicherheitstraining zu evaluieren.

Wer führt die Untersuchung durch?

Die Untersuchung wird durchgeführt von Ryo Toyoda, unter der Supervision von Prof. Jarka Glassey, Dr. Fernando Russo Abegão, und Dr. Sue Gill of Newcastle University, als Teil des Europäischen Trainingsnetzwerk für Chemical Engineering Immersives Lernen (ETN-CHARMING) Projekt.

Wie genau nehme ich teil?

Sie sind eingeladen an einer Onlineumfrage teilzunehmen. Die Umfrage sollte weniger als 15 Minuten Ihrer Zeit in Anspruch nehmen. Die Fragen beziehen sich auf Ihr Empfinden über die Verwendung des VR-Spiels für das G&S-Mitarbeiter-Training in der Chemieindustire.

Was passiert mit meinen Antworten?

Bitte seien Sie versichert, dass Ihre Antworten vollständig **anonym** und **vertraulich** sind. Nur Personen, die direkt an dieser Umfrage beteiligt sind, haben Zugriff auf die von uns gesammelten Rohdaten. Es besteht kein Risiko, an dieser Umfrage teilzunehmen. Die Informationen werden sicher in einem geschützten Dateispeicher gespeichert. Das Umfrageergebnis und der Bericht enthalten keine Angaben zur Person. Die anonymisierten Daten und die daraus resultierende Analyse können in akademischen und nichtakademischen Veröffentlichungen und Präsentationen verwendet werden.

Was ist wenn ich meine Meinung ändere oder Fragen habe?

Sie können jederzeit Fragen stellen, die Teilnahme stoppen oder auch komplett zurückziehen.

Nachdem Sie diese Information gelesen und sich für die Teilnahme am Projekt entschieden haben, werden Sie am Beginn der Umfrage gebeten werden, die beigefügte Einverständniserklärung auszufüllen.

Falls Sie Fragen haben, bitte zögern Sie nicht Kontakt aufzunehmen mit:

Ryo Toyoda
ryo.toyoda@newcastle.ac.uk
PhD | School of Engineering | Newcastle University
Merz Court | Newcastle upon Tyne | NE1 7RU | UK

Mitarbeiter-Einverständniserklärung für die Forschungsumfrage als Teil des Gesundheits- und Sicherheitstraining (G&S) mit Virtual Reality (VR) Spiel Projekts

Ich bestätige, dass:

- Ich die Informationen zum Projekt, die am Informationsblatt bereitgestellt sind, gelesen und verstanden habe.
- Ich die Möglichkeit hatte, Fragen zum Projekt und zu meiner Teilnahme zu stellen.
- Ich freiwillig teilnehme
- Ich verstehe, dass ich meine Teilnahme am Projekt jederzeit ohne Angabe von Gründen zurückziehen kann und ich nicht darüber befragt werde
- Ich verstehe, dass andere Forscher der Newcastle University und vom CHARMING Projekt Zugang zu den anonymisierten Informationen haben und dass die Forscher dem am Informationsblatt beschriebenen Prozedere für die Datenspeicherung folgen werden.
- Ich verstehe, dass ich eine Onlineumfrage für diese Studie beantworten werde und dass meine Antworten sicher auf einem gesicherten Dateispeicher gespeichert werden
- Ich verstehe, dass anonymisierte Antworten aus der Umfrage in schriftlichen akademischen und nicht-akademischen Veröffentlichungen und Konferenzpräsentationen zum Forschungsprojekt verwendet werden können.

☐ Ich habe die AGB gelesen und bin damit einverstanden. (Bitte ankreuzen)

Teil A: Demografische Daten

Firma: _____________

Job Beschreibung: _____________

Geschlecht

- Weiblich
- Männlich

Was ist Ihre Altersgruppe

- Unter 20
- 20-29
- 30-39
- 40-49
- 50-59
- 60& darüber

Herkunftsland: ______________

Was ist Ihre höchste abgeschlossene Ausbildung?

- Grundschule
- Sekundarstufe I
- Sekundarstufe II
- Bachelor Studium
- Master Studium
- Doktoratsstudium
- Andere: ________

Wie lange arbeiten Sie bereits im Unternehmen?

- Weniger als ein Jahr
- 1-5 Jahre
- 6-20 Jahre
- Mehr als 20 Jahre

Welche Form des G&S-Trainings gibt es in Ihrem Unternehmen? (Sie können mehr als eine Antwort auswählen)

- Formeller Unterricht (Vortrag)
- Informeller Unterricht (zB kleine Gruppendiskussionen, Fallbeispiele)
- Coaching oder Ausbildung am Arbeitsplatz
- Offenes Lernen und Fernunterricht (zB Video Präsentationen, Online Plattform)
- Virtual Reality (VR) und/oder Augmented Reality (AR)
- Andere: _________

Teil B: Virtual Reality (VR) / Spielerfahrung

Haben Sie schon einmal versucht ein (am Kopf montiertes) Virtual Reality Display zu verwenden?

- Ja
- Nein

Welche VR-Spielekonsole oder ähnliches haben Sie verwendet? (Sie können mehrere Antworten auswählen)

- Sony PlayStation VR
- HTC Vive
- Oculus Rift/quest
- Samsung Gear VR
- Google Cardboard
- LG 360 VR
- Google Daydream
- Keine
- Andere: ________

Wie oft in einem Jahr haben Sie diese verwendet?

- Nie
- Einmal
- 2-5 mal
- Mehr als 5 mal

Spielen Sie Videospiele?

- Ja
- Nein

Welches Spielkonsole oder ähnliches haben Sie verwendet? (Sie können mehrere Antworten auswählen)

- Nintendo (Gameboy, 3DS, WII, Switch, usw)
- PlayStation
- Xbox
- Handyspiele
- Computerspiele
- Keine
- Andere: _______

Wie oft in der Woche spielen Sie Videospiele?

- Nie
- Weniger als 5 Stunden
- 5-15 Stunden
- 16-25 Stunden
- Mehr als 25 Stunden

Teil C: Dieser Teil versucht zu verstehen, wie Sie den Einsatz vom Virtual Reality (VR) Spiel für die Mitarbeiter G&S-Schulung in der Chemieindustrie empfinden

Items	6-point Likert Scale					
	Stimme stark zu	Stimme zu	Stimme eher zu	Stimme eher nicht zu	Stimme nicht zu	Lehne stark ab
1. Ich denke, dass die Verwendung der VR-Umgebung nützlich sein						

wird, um G&S-Verfahren zu üben.						
2. Die Verwendung einer VR-Umgebung wird mir wahrscheinlich helfen, die G&S-Verfahren schneller zu lernen.						
3. Wenn ich diese VR-Umgebung verwende, werde ich meine Leistung bei G&S-Verfahren verbessern.						
4. Ich denke, dass die Verwendung der VR-Umgebung klar und verständlich sein wird.						
5. Ich denke, dass es für mich einfach sein wird, die Plattform zu bedienen, auf der die VR-Umgebung ausgeführt wird.						
6. Ich denke, dass es zu lange dauern wird, um zu lernen wie man die VR-Umgebung verwendet, als dass sich die Mühe lohnt.						
7. Ich denke, dass die Organisation mich darin unterstützen wird, den Umgang mit der VR-Umgebung zu lernen.						
8. Leute, die mein Verhalten bei der Arbeit beeinflussen, denken, dass ich diese VR-Umgebung verwenden sollte.						
9. Ich denke, dass mein Vorgesetzter die Nutzung dieser VR-Umgebung in meiner Arbeit sehr unterstützen wird.						
10. Ich fühle, dass es eine schlechte Idee ist, die VR-Umgebung für die G & H-Schulung zu verwenden.						
11. Ich denke, dass der eigentliche Nutzungsprozess der VR-Umgebung für das G & H-Schulung Spaß macht.						

12. Ich denke, dass es sehr frustrierend sein wird, die VR-Umgebung für das G & H-Schulung einzusetzen.						
13. Wenn es mir zur Verfügung gestellt wird, würde ich die Verwendung der VR-Umgebung für das Anwendungslernen von G&H-Verfahren bei meinen KollegInnen empfehlen.						
14. Wenn es mir zur Verfügung gestellt wird, plane ich, die VR-Umgebung weiterhin häufig für G & H-Schulungen zu verwenden.						
15. Ich denke, dass ich nach der Verwendung der VR für die G & H-Schulung bereit sein werde, diese Lernumgebung bei einem anderen Schulungskurs zu verwenden.						

Appendix C – Online questionnaire used in the UTAUT2 study (French)

Formation H&S utilisant des jeux de Réalité Virtuelle (RV)

Quel est l'objectif de ce questionnaire?

Le but du projet de recherche est d'évaluer l'acceptation de la nouvelle technologie de jeu en réalité virtuelle (RV) dans le domaine de la formation H&S, ainsi que l'intention d'utiliser cette nouvelle technologie.

Qui conduit ce projet de recherche?

Cette recherche est menée par Ryo Toyoda sous la supervision du Pr. Jarka Glassey, du Dr. Fernando Russo Abegão, et du Dr. Sue Gill de l'université de Newcastle, dans le cadre du projet de réseau européen de formation pour l'apprentissage immersif en génie chimique (projet ETN-CHARMING).

En quoi consiste ma participation?

Vous êtes invité à répondre à un questionnaire en ligne. Cela devrait vous prendre moins de 15 minutes. Les questions concerneront votre perception de l'usage de jeu en réalité virtuelle (RV) pour la formation HSE des salariés de l'industrie chimique.

Qu'allez-vous faire de mes réponses au questionnaire?

Toutes les réponses que vous faites sont complètement anonymes et confidentielles. Seules les personnes directement impliquées dans cette recherche ont accès aux données que nous collectons. La participation à cette étude ne présente aucun risque. Ces informations sont stockées de façons sécurisées dans un service de stockage de fichiers protégés. Les résultats de recherche et les rapports ne feront pas référence à un individu particulier. Les données anonymisées et les analyses résultantes pourront être utilisées dans des publications et présentations académiques ou non académiques.

Puis-je changer d'avis ou poser des questions?

A n'importe quel moment vous pouvez poser des questions, arrêter votre participation ou vous retirer complètement du projet de recherche.
Après avoir lu ces informations, et si vous êtes d'accord pour prendre part au projet, vous devrez compléter un formulaire de consentement au début du questionnaire.

Si vous avez des questions, n'hésitez pas à contacter:

Ryo Toyoda
ryo.toyoda@newcastle.ac.uk
Doctorant | School of Engineering | Newcastle University
Merz Court | Newcastle upon Tyne | NE1 7RU | UK

Formulaire de consentement au questionnaire concernant le projet de recherche de formation H&S utilisant des jeux de réalité virtuelle

Je confirme que:

- J’ai lu et compris les informations concernant le projet, tel que décrit dans la fiche d’information.
- J’ai eu la possibilité de poser des questions concernant le projet et ma participation au projet.
- Je suis entièrement d’accord pour y participer.
- Je comprends que je peux arrêter ma participation au questionnaire à tout moment, sans qu’on me pose de questions.
- Je comprends que d’autres chercheurs de l’Université de Newcastle et du consortium CHARMING faisant partie de l’équipe projet pourront avoir accès aux données anonymes qui auront été collectées et que ces chercheurs respecteront les procédures de stockage de données telles que décrites dans la fiche d’information.
- Je comprends que je vais répondre à un questionnaire en ligne pour cette étude et que mes réponses seront stockées de façons sécurisées dans un service de stockage de fichiers protégés.
- Je comprends que les réponses anonymisées au questionnaire pourront être utilisées dans des publications écrites académiques ou non académiques et dans des présentations en conférence à propos du projet de recherche.

☐ J’ai lu et accepte les termes et conditions. (Merci de cocher la case)

Partie A : profil

Entreprise: ____________

Fonction dans l'entreprise: ____________

Sexe :
- Femme
- Homme

Quel âge avez-vous?
- Moins de 20 ans
- 20-29
- 30-39
- 40-49
- 50-59
- plus de 60 ans

Pays de naissance or Nationalité? _____________

Niveau de formation le plus élevé (ou équivalent)?
- École primaire
- Collège
- Lycée
- Licence / bac +2 ou 3
- Master / bac +4 ou 5
- Doctorat
- Autre: ______________

Depuis combien de temps travaillez-vous dans cette entreprise?
- Moins d'un an
- 1-5 ans
- 6-20 ans
- Plus de 20 ans

Quels types de formations HSE sont disponibles dans votre entreprise? (plusieurs choix possible)
- Dans un cadre formel en salle (comme un cours)
- Dans un cadre informel en salle (par exemple petit groupe de discussion, étude de cas)
- Coaching/compagnonnage ou formation sur le terrain
- Enseignement ouvert et à distance (par exemple vidéo ou e-learning)
- Dans un environnement de réalité virtuelle (RV) ou réalité augmentée (RA)
- Autres: _______________

Partie B : Expérience en réalité virtuelle (RV) / jeu

Avez-vous déjà essayé un casque de réalité virtuelle?
- Oui
- Non

Quelle(s) console(s) de jeu en RV ou autres dispositifs de RV avez-vous déjà utilisé? (plusieurs réponses possible)

- Sony PlayStation RV
- HTC Vive
- Oculus Rift/quest
- Samsung Gear RV
- Google Cardboard
- LG 360 RV
- Google Daydream
- Aucune
- Autres: ________

Combien de fois par an l'utilisez-vous?

- Jamais
- Une fois
- 2-5 fois
- Plus de 5 fois

Jouez-vous à des jeux vidéo?

- Oui
- Non

Quel dispositif de jeu vidéo avez-vous utilisé? (plusieurs réponses possible)

- Nintendo (Gameboy, 3DS, WII, Switch, *etc.*)
- PlayStation
- Xbox
- Jeux sur smartphone
- Jeux sur ordinateur
- Aucun
- Autres: ________

Combien de temps jouez-vous par semaine ?

- Jamais
- Moins de 5 heures
- 5-15 heures
- 16-25 heures
- Plus de 25 heures

Partie C : cette section vise à comprendre votre perception de l'usage des jeux en réalité virtuelle (RV) pour la formation H&S dans l'industrie chimique

	6-point Likert Scale					
	Je suis totalement d'accord	Je suis d'accord	Je suis assez d'accord	Je ne suis pas tout à fait d'accord	Je ne suis pas d'accord	Je ne suis vraiment pas d'accord
1. Je pense que l'utilisation d'un						

univers de RV sera utile pour s’entrainer sur les procédures H&S.						
2. L’utilisation de l’univers RV me permettra sans doute d’apprendre les procédures H&S plus rapidement.						
3. Si j’utilise l’univers RV, j’améliorerai mes performances sur les procédures H&S.						
4. Je pense que l’univers RV sera clair et compréhensible.						
5. Je pense que ce sera simple pour moi de me servir de l’outil sur lequel fonctionne l’univers de réalité virtuelle.						
6. Je pense que cela prendra trop de temps pour apprendre à utiliser l’univers RV pour que cela vaille la peine de s’y mettre.						
7. Je pense que l’entreprise m’aidera à apprendre à utiliser l’univers RV.						
8. Les personnes qui influencent mon comportement au travail pense que je devrais utiliser l’environnement RV.						
9. Je pense que mon chef sera très favorable à l’utilisation de						

l’univers RV pour mon poste de travail.						
10. Je pense que c’est une mauvaise idée d’utiliser un univers RV pour la formation H&S.						
11. Je pense que l’utilisation d’un univers RV pour la formation H&S est fun.						
12. Je pense que l’utilisation d’un univers RV pour la formation H&S sera exaspérante.						
13. Si c’était disponible, je recommanderais à mes collègues l’utilisation d’un univers RV pour apprendre comment appliquer les procédures H&S.						
14. Si c’était disponible, j’utiliserais fréquemment l’univers RV pour les formations H&S.						
15. Je pense qu’après avoir utilisé l’environnement RV pour la formation H&S, je serai prêt à utiliser cet environnement de formation dans d’autres parcours de formation.						

Appendix D – Online questionnaire used in the IVR-based assessment study (English)

Research project information sheet 2021

What is the purpose of the research?

The purpose of this experiment is to test a virtual reality prototype. This prototype was developed to evaluate virtual reality training as part of chemical operator training at Merck. We will ask you for feedback on your experience. Another goal of the study is to examine the user experience during the training process.

Who researches?

This research is carried out by Sofia Garcia Fracaro, Yusra Tehreem, Tim Gallagher and Ryo Toyoda under the supervision of Dr Michael Wilk (Merck), Prof Thies Pfeiffer (Technik Emden), Prof Jarka Glassey (University of Newcastle), Prof Kristel Bernaerts (KU Leuven), as part of the European Training Network for Chemical Engineering Immersive Learning (ETN-CHARMING) project.

How do I specifically participate?

You will be invited to test a Virtual Reality prototype. Your participation also includes a survey, before and after the Virtual Reality experience.

What happens to the information you have collected about me?

Only people directly involved in this research have access to the data we collect. The information is stored securely in the protected file storage service. Please be assured that your responses are completely anonymous and confidential. The research result and the report do not contain any reference to an individual. The anonymised data and resulting analysis may be used in academic and non-academic publications and presentations.

Are there any risks?

There is no risk involved in participating in this study. Your results, experiences and evaluations will be handled in such a way that no personal conclusions can be drawn.

What if I change my mind or have questions?

You can ask questions, stop participating or withdraw completely from the research project at any time.

If you have any questions, please contact:
Sofia Garcia Fracaro - Sofia.garcia-fracaro@merckgroup.com

Declaration of consent for a research survey as part of the research project

I, the undersigned, certify the following:

- I have read and understood the information on the project according to the information sheet.
- I had the opportunity to ask questions about the project and my participation in it.
- I voluntarily agree to my participation.
- I understand that I may withdraw my participation at any time without giving any reason and that I will not be penalised for withdrawing nor asked why I have withdrawn.
- I understand that other researchers in the CHARMING consortium who are part of the project team, may have access to the information collected and that the researchers will follow the agreed procedures for the storage of the data, which are set out in the information sheet.
- I understand that I am answering surveys for the study and the answers will be stored securely in the protected file storage service.
- I understand that anonymised responses from the survey may be used in written academic and non-academic publications and conference presentations on the research project.
- I consent to photographs being taken of my participation. I understand that anonymised photographs may be used in written academic and non-academic publications and conference presentations as well as at science communication events.

I have read the general terms and conditions and agree to them.

- Yes
- No

Pre-Questionnaire

Note on data protection in the survey: Participation in the survey is voluntary. If you decide to participate, you will remain anonymous. If you do not wish to participate, there will be no negative consequences for you. To avoid drawing conclusions about the identity of survey participants, we only evaluate questions that have received at least five responses. Please do not provide information about your name or the names of other persons.

Demographic profile

1. Participant number: ___________

2. What is your age group?
   - Less than 20 years old
   - 20-30 years old
   - 31-40 years old
   - 41-50 years old
   - 51-60 years old
   - Over 60 years old
3. What is your gender?
   - Female
   - Male
   - Prefer not to say

Post-Questionnaire

Use of the assessment overview syestem (Step 3)

This section examines how you assess the use of the assessment overview system (step 3) in virtual reality (VR).

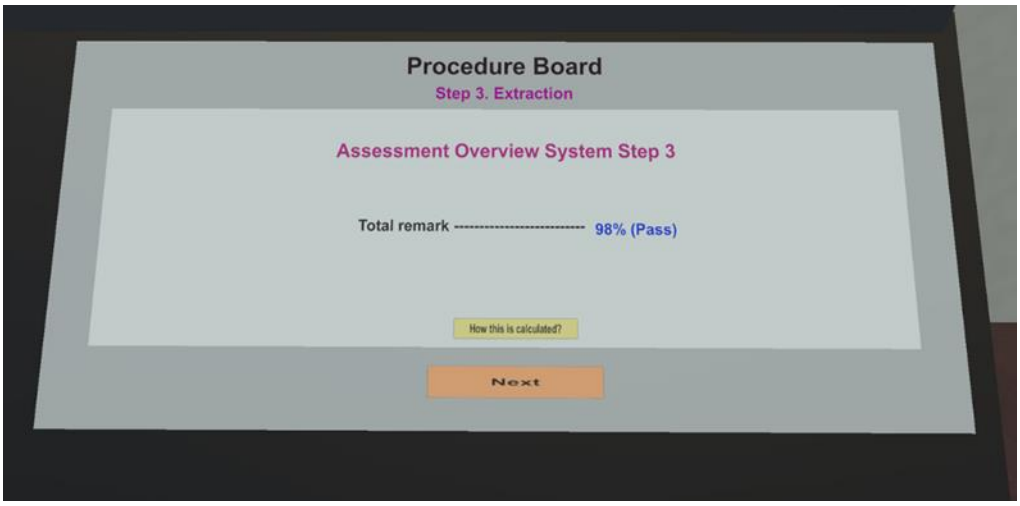


Photo shown for control group during the survey.

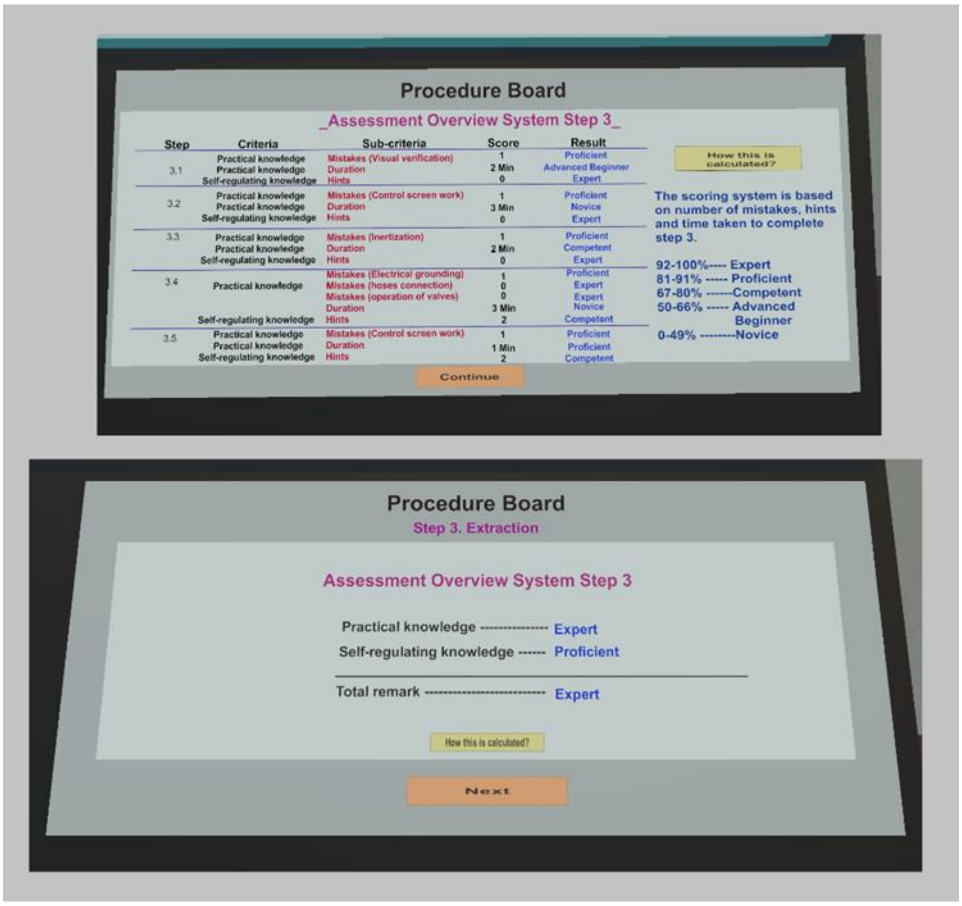


Photo shown for experimental group during the survery.

Please answer the following questions:

	6-point Likert Scale					
	Strongly Agree	Agree	Slightly Agree	Slightly Disagree	Disagree	Strongly Disagree
1. Using the assessment overview system is favourable.						
2. Using the assessment overview system is a bad idea.						
3. Using the assessment overview system is beneficial.						
4. Using the assessment overview system is foolish.						
5. Assuming I had access to the assessment						

overview system, I intend to use it.						
6. Given that I had access to the assessment overview system, I predict that I would use it.						
7. I plan to use the assessment overview system in the future.						

Please answer the following questions:

	6-point Likert Scale					
	1 (Never)	2	3	4	5	6 (Always)
1. Did the assessment overview system provide the precise information you need?						
2. Did the assessment overview system provide reports that seem to be just about exactly what you need?						
3. Did the assessment overview system provide sufficient information?						

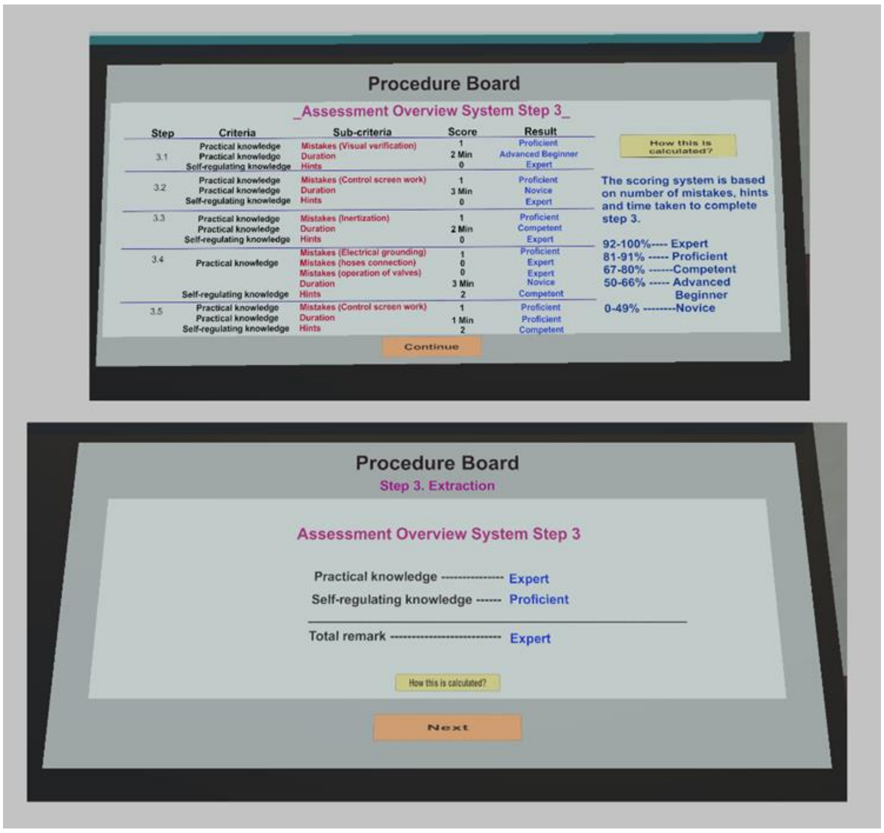
What do you think are the advantages of using the assessment overview system?

____
____
____

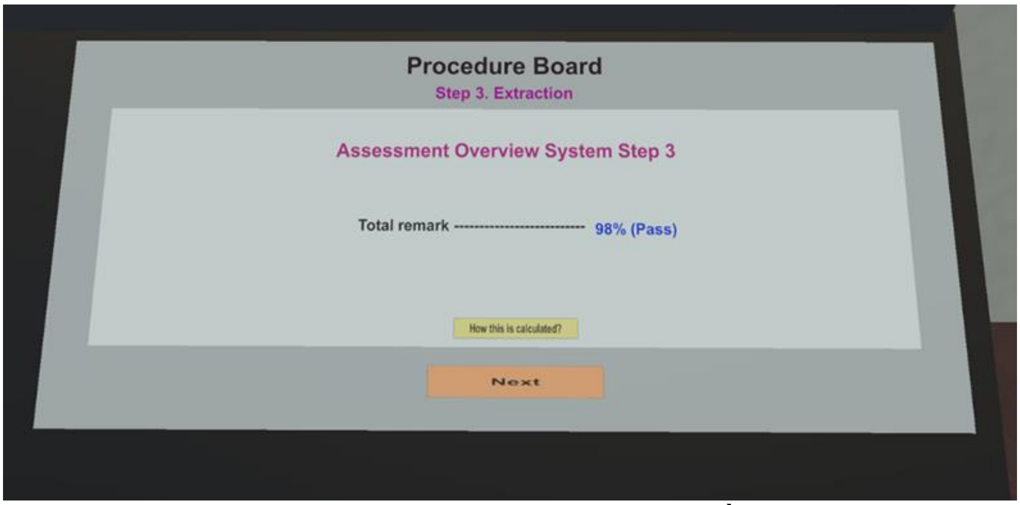
What do you think are the disadvantages of using the assessment overview system?

____
____
____
____

Below is the picture of the detailed assessment overview system (1st photo) and the simple assessment overview system (2nd photo)



Detailed assessment overview system (1st photo)



Simple assessment overview system (2nd photo)

Which assessment overview system do you prefer in the VR prototype? Kindly choose from the choices below.

- I prefer the detailed version (1st photo)
- I prefer the version I have seen in the VR prototype (2nd photo)

Kindly explain why do you prefer the said assessment overview system.

____

____

____

Appendix E – Online questionnaire used in the IVR-based assessment study (German)

Forschungsprojekt Informationsblatt 2021

Was ist der Zweck der Forschung?

Das Ziel dieser Erfahrung ist es, einen Virtual Reality-Prototyp zu testen. Dieser Prototyp wurde entwickelt, um das Virtual-Reality-Training als Teil der Ausbildung von Chemiebedienern bei Merck zu evaluieren. Wir werden Sie um ein Feedback zu Ihrer Erfahrung bitten. Ein weiteres Ziel der Studie ist es, Ihre Benutzererfahrung während des Trainingsvorgangs zu erforschen.

Wer forscht?

Diese Forschung wird von Sofia Garcia Fracaro, Yusra Tehreem, Tim Gallagher und Ryo Toyoda unter der Leitung von Dr. Michael Wilk (Merck), Prof. Thies Pfeiffer (Technik Emden), Prof. Jarka Glassey (University of Newcastle), Prof. Kristel Bernaerts (KU Leuven), als Teil des European Training Network for Chemical Engineering Immersive Learning (ETN-CHARMING) Projekts durchgeführt.

Wie nehme ich konkret teil?

Sie werden eingeladen, einen Virtual Reality-Prototyp zu testen. Ihre Teilnahme beinhaltet auch eine Umfrage, vor und nach der Virtual Reality-Erfahrung.

Was passiert mit den Informationen, die Sie über mich gesammelt haben?

Nur Personen, die direkt an dieser Untersuchung beteiligt sind, haben Zugriff auf die von uns gesammelten Daten. Die Informationen werden sicher im geschützten Dateispeicherdienst gespeichert. Bitte seien Sie versichert, dass Ihre Antworten vollständig anonym und vertraulich sind. Das Forschungsergebnis und der Bericht enthalten keinen Verweis auf eine Person. Die anonymisierten Daten und die daraus resultierende Analyse können in akademischen und nichtakademischen Veröffentlichungen und Präsentationen verwendet werden.

Gibt es irgendwelche Risike?

Die Teilnahme an dieser Studie birgt kein Risiko. Ihre Ergebnisse, Erfahrungen und Bewertungen werden so gehandhabt, dass keinerlei personenbezogene Schlussfolgerungen gezogen werden können.

Was ist, wenn ich meine Meinung ändere oder Fragen habe?

Sie können jederzeit Fragen stellen, Ihre Teilnahme einstellen oder sich vollständig aus dem Forschungsprojekt zurückziehen.

Bei Fragen wenden Sie sich bitte an:
Sofia Garcia Fracaro - Sofia.garcia-fracaro@merckgroup.com

Einverständniserklärung für eine Forschungsumfrage im Rahmen des Forschungsprojekts

Ich, der Unterzeichnete, bestätige Folgendes:

- Ich habe die Informationen zum Projekt gemäß dem Informationsblatt gelesen und verstanden.
- Ich hatte die Möglichkeit, Fragen zum Projekt und meiner Teilnahme daran zu stellen.
- Ich stimme meiner Teilnahme freiwillig zu.
- Ich verstehe, dass ich meine Teilnahme jederzeit ohne Angabe von Gründen widerrufen kann und dass ich weder für den Widerruf bestraft noch gefragt werde, warum ich zurückgetreten bin.
- Ich verstehe, dass andere Forscher des CHARMING-Konsortiums, die Teil des Projektteams sind, möglicherweise Zugriff auf die gesammelten Informationen haben und dass die Forscher die vereinbarten Verfahren für die Speicherung der Daten befolgen, die im Informationsblatt aufgeführt sind.
- Ich verstehe, dass ich Umfragen für die Studie beantworte und die Antworten sicher im geschützten Dateispeicherdienst gespeichert werden.
- Ich verstehe, dass anonymisierte Antworten aus der Umfrage in schriftlichen akademischen und nichtakademischen Veröffentlichungen und Konferenzpräsentationen zum Forschungsprojekt verwendet werden können.
- Ich bin damit einverstanden, dass Fotos von meiner Teilnahme gemacht werden. Mir ist bekannt, dass anonymisierte Fotos in schriftlichen akademischen und nicht-akademischen Publikationen und Konferenzpräsentationen sowie bei Veranstaltungen zur Wissenschaftskommunikation verwendet werden können.

Ich habe die Allgemeinen Geschäftsbedingungen gelesen und bin damit einverstanden.

- Ja
- Nein

Fragebogen vor der virtuellen Realität

Hinweis zum Datenschutz bei der Umfrage: Die Teilnahme an der Umfrage ist freiwillig. Wenn Sie sich für die Teilnahme entscheiden, bleiben Sie anonym. Wenn Sie nicht teilnehmen möchten, hat das keine negativen Auswirkungen für Sie. Um Rückschlüsse auf die Identität der Umfrageteilnehmer zu vermeiden, werten wir nur Fragen aus, auf die mindestens fünf Antworten eingegangen sind. Bitte machen Sie keine Angaben zu Ihrem Namen oder den Namen anderer Personen.

Demographisches Profil

1. Teilnehmernummer: ___________

2. Was ist ihre Altersgruppe?
    - Weniger als 20 Jahre alt
    - 20-30 Jahre alt
    - 31-40 Jahre alt
    - 41-50 Jahre alt
    - 51-60 Jahre alt
    - Über 60 Jahre alt

3. Was ist ihr Geschlecht?
    - Weiblich
    - Männlich
    - Möchte ich nicht sagen

Fragebogen nach der virtuellen Realität

Verwendung des Bewertungsübersichtssystems (Stufe 3)

Dieser Teil versucht zu verstehen, wie Sie die Verwendung des Bewertungsübersichtssystems (Stufe 3) in der virtuellen Realität (VR) einschätzen.

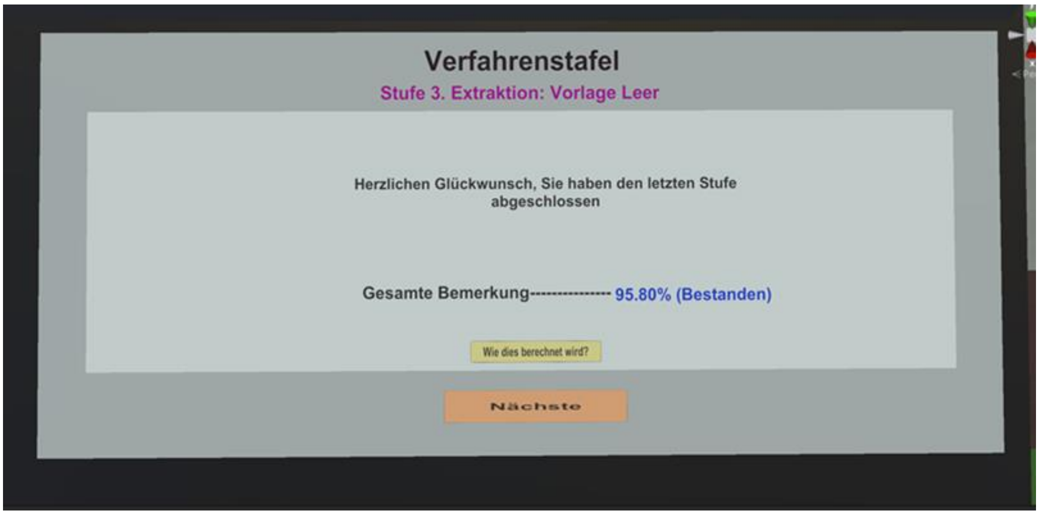


Photo shown for control group during the survey.

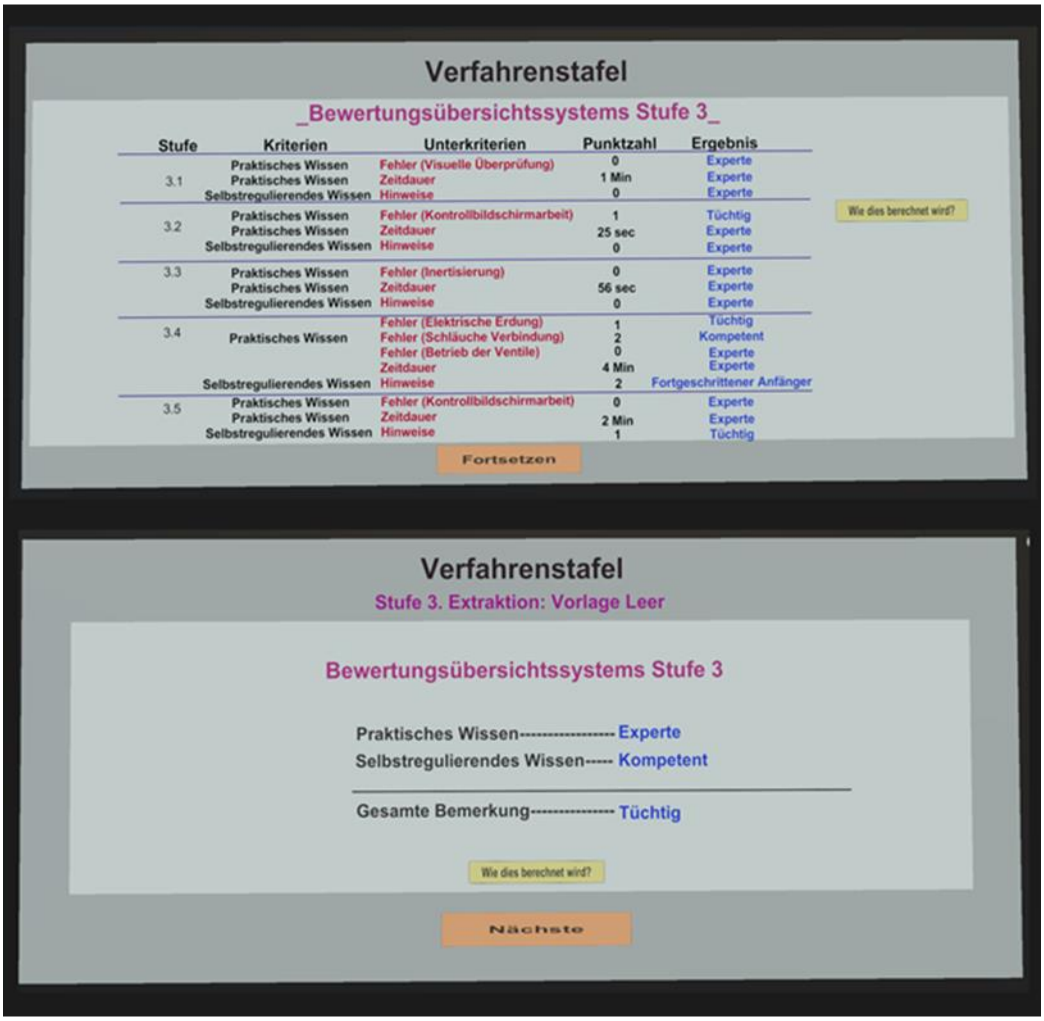


Photo shown for experimental group during the survery.

Bitte beantworten Sie die folgenden Fragen:

	6-point Likert Scale					
	Stimme stark zu	Stimme zu	Stimme eher zu	Stimme eher nicht zu	Stimme nicht zu	Lehne stark ab
1. Die Verwendung des Bewertungsübersichtssystems ist günstig.						
2. Die Verwendung des Bewertungsübersichtssystems ist eine schlechte Idee.						
3. Die Verwendung des Bewertungsübersichtssystems ist vorteilhaft.						
4. Die Verwendung des Bewertungsübersichtssystems ist töricht.						
5. Angenommen, ich hätte Zugang zum Bewertungsübersichtssyste						

ms, dann würde ich es nutzen.						
6. Wenn ich Zugang zum Bewertungsübersichtssystems hätte, würde ich es voraussichtlich nutzen.						
7. Ich plane, das Bewertungsübersichtssystems in Zukunft zu nutzen.						

Bitte beantworten Sie die folgenden Fragen:

	6-point Likert Scale					
	1 (Nie)	2	3	4	5	6 (Immer)
1. Hat das Bewertungsübersichtssystems genau die Informationen geliefert, die Sie benötigen?						
2. Hat das Bewertungsübersichtssystems Berichte geliefert, die genau das zu sein scheinen, was Sie brauchen?						
3. Hat das Bewertungsübersichtssystems ausreichende Informationen geliefert?						

Was sind Ihrer Meinung nach die Vorteile des Bewertungsübersichtssystems?

___

___

___

Was sind Ihrer Meinung nach die Nachteile des Bewertungsübersichtssystems?

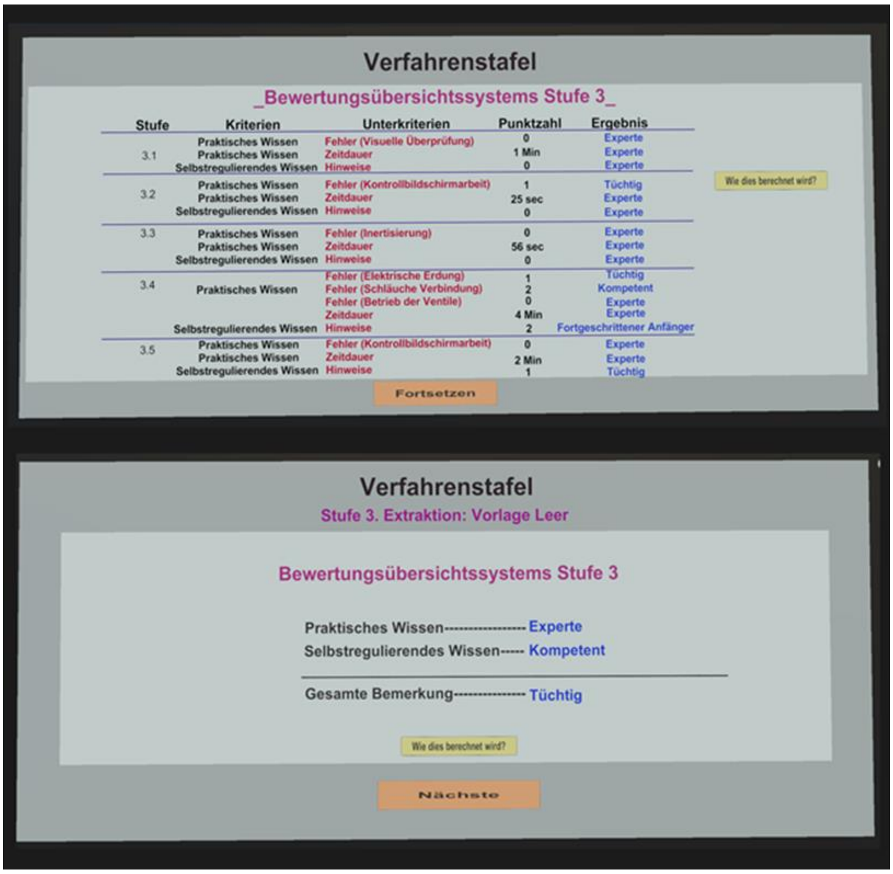
___

___

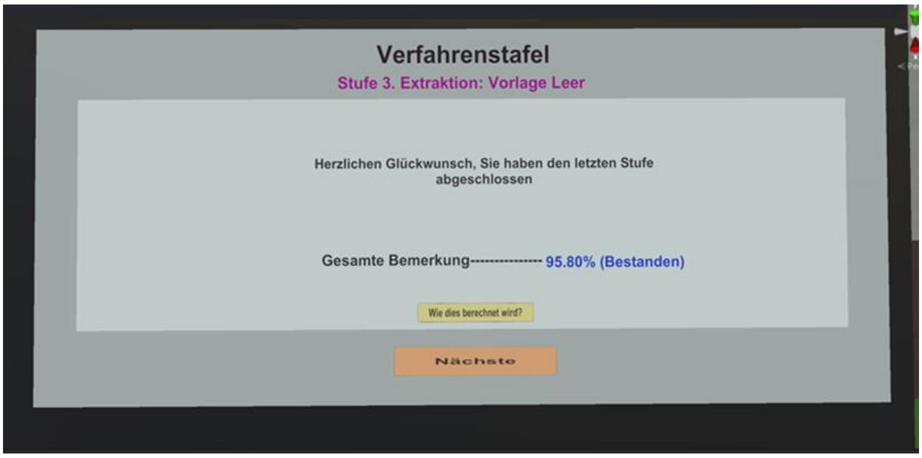
___

___

Hieronder vindt u de foto van het gedetailleerde overzichtssysteem voor beoordeling (1e foto) en het simpele overzichtssysteem voor beoordeling (2e foto)



Unten sehen Sie das Foto des detaillierten Bewertungsübersichtssystems (1e foto)



Unten sehen Sie das Foto des einfacheBewertungsübersichtssystems. (2e foto)

Welches Bewertungsübersichtssystems bevorzugen in der virtuellen Realität (VR)? Bitte wählen Sie aus den unten stehenden Möglichkeiten.

- Ich bevorzuge die detaillierte Version (1. Foto)
- Ich bevorzuge die einfache Version (2. Foto)

Erläutern Sie bitte, warum Sie dieses Bewertungsübersichtssystems bevorzugen.

____

____

____

Appendix F – Online questionnaire used in the IVR-based assessment study (Dutch)

Informatieblad onderzoeksproject 2021

Wat is het doel van het onderzoek?

Het doel van deze ervaring is het testen van een virtual reality-prototype. Dit prototype is ontwikkeld om virtual reality training te evalueren als onderdeel van de opleiding van chemisch operators bij Merck. Wij zullen u vragen om feedback over uw ervaring. Een ander doel van de studie is de gebruikerservaring tijdens het opleidingsproces te onderzoeken.

Wie doet het onderzoek?

Dit onderzoek wordt uitgevoerd door Sofia Garcia Fracaro, Yusra Tehreem, Tim Gallagher en Ryo Toyoda onder leiding van Dr. Michael Wilk (Merck), Prof. Thies Pfeiffer (Technik Emden), Prof. Jarka Glassey (University of Newcastle), Prof. Kristel Bernaerts (KU Leuven), als onderdeel van het European Training Network for Chemical Engineering Immersive Learning (ETN-CHARMING) project.

Hoe kan ik concreet deelnemen?

U zult worden uitgenodigd om een Virtual Reality-prototype te testen. Uw deelname omvat ook een enquête, voor en na de Virtual Reality ervaring.

Wat gebeurt er met de informatie die u over mij hebt verzameld?

Alleen personen die rechtstreeks bij dit onderzoek betrokken zijn, hebben toegang tot de informatie die wij verzamelen. De informatie wordt veilig opgeslagen in de beveiligde bestandsopslagdienst. U kunt er zeker van zijn dat uw antwoorden volledig anoniem en vertrouwelijk zijn. Het onderzoeksresultaat en het verslag bevatten geen enkele verwijzing naar een persoon. De geanonimiseerde gegevens en de daaruit voortvloeiende analyse mogen worden gebruikt in academische en niet-academische publicaties en presentaties.

Zijn er risico's?

Er is geen risico aan deelname aan deze studie. Uw resultaten, ervaringen en beoordelingen zullen zo worden behandeld dat er geen persoonlijke conclusies uit kunnen worden getrokken.

Wat als ik van gedachten verander of vragen heb?

U kunt op elk moment vragen stellen, uw deelname stopzetten of u volledig terugtrekken uit het onderzoeksproject.

Als u vragen heeft, kunt u contact opnemen met:
Sofia Garcia Fracaro - Sofia.garcia-fracaro@merckgroup.com

Toestemmingsverklaring voor een onderzoeksenquête in het kader van het onderzoeksproject

Ondergetekende bevestigt het volgende:

- Ik heb de informatie over het project volgens het informatieblad gelezen en begrepen.
- Ik heb de gelegenheid gehad om vragen te stellen over het project en mijn deelname eraan.
- Ik ga vrijwillig akkoord met mijn deelname.
- Ik begrijp dat ik mijn deelname op elk moment zonder opgaaf van redenen kan intrekken en dat ik niet zal worden gestraft voor mijn terugtrekking noch zal worden gevraagd waarom ik mij heb teruggetrokken.
- Ik begrijp dat andere onderzoekers van het CHARMING-consortium die deel uitmaken van het projectteam, toegang kunnen hebben tot de verzamelde informatie en dat de onderzoekers de overeengekomen procedures voor de opslag van de gegevens zullen volgen, zoals uiteengezet in het informatieblad.
- Ik begrijp dat ik enquêtes beantwoord voor het onderzoek en dat de antwoorden veilig zullen worden opgeslagen in de beveiligde bestandsopslagdienst.
- Ik begrijp dat geanonimiseerde antwoorden op de enquête kunnen worden gebruikt in schriftelijke academische en niet-academische publicaties en presentaties op conferenties over het onderzoeksproject.
- Ik geef toestemming om foto's te maken van mijn deelname. Ik begrijp dat geanonimiseerde foto's kunnen worden gebruikt in schriftelijke academische en niet-academische publicaties en presentaties op conferenties en bij wetenschapscommunicatie-evenementen.

Ik heb de algemene voorwaarden gelezen en ga ermee akkoord.

- Ja
- Nee

Vragenlijst vóór virtuele realiteit

Opmerking over gegevensbescherming in de enquête: Deelname aan de enquête is vrijwillig. Als u besluit deel te nemen, blijft u anoniem. Indien u niet wenst deel te nemen, zal dit geen negatieve gevolgen voor u hebben. Om te voorkomen dat er conclusies worden getrokken over de identiteit van de deelnemers aan de enquête, evalueren wij alleen vragen waarop ten minste vijf reacties zijn binnengekomen. Gelieve geen informatie te verstrekken over uw naam of de namen van andere personen.

Demografisch profiel

1. Deelnemersnummer: ___________

2. Wat is hun leeftijdsgroep?
   - Minder dan 20 jaar oud
   - 20-30 jaar oud
   - 31-40 jaar oud
   - 41-50 jaar oud
   - 51-60 jaar oud
   - Meer dan 60 jaar oud

3. Wat is uw geslacht?
   - Vrouw
   - Mannelijk
   - Zeg liever niet

Vragenlijst na de virtuele realiteit

Gebruik van het overzichtssysteem voor beoordeling (Stap 3)

In dit deel wordt nagegaan hoe u het gebruik van het overzichtssysteem voor beoordeling (stap 3) in virtual reality (VR) beoordeelt.

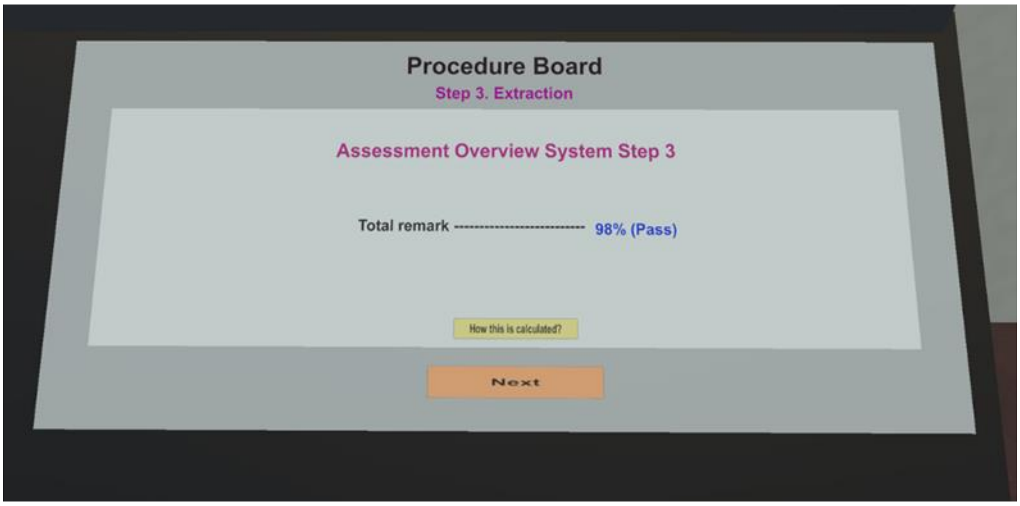


Photo shown for control group during the survey.

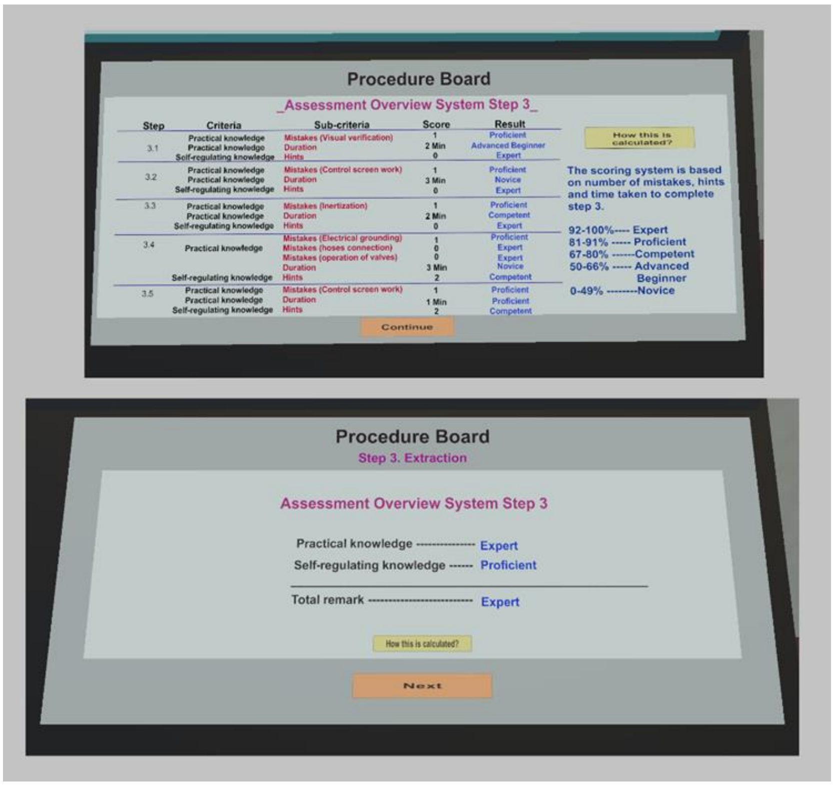


Photo shown for experimental group during the survery.

Gelieve de volgende vragen te beantwoorden:

	6-point Likert Scale					
	helemaal mee eens	mee eens	een beetje mee eens	een beetje mee oneens	mee oneens	helemaal mee oneens
1. Het gebruik van het overzichtssysteem voor beoordeling is gunstig.						
2. Het gebruik van het overzichtssysteem voor beoordeling is een slecht idee.						
3. Het gebruik van het overzichtssysteem voor beoordeling is nuttig.						
4. Het gebruik van het overzichtssysteem voor beoordeling is dom.						

5. Als ik toegang had tot het overzichtssysteem voor beoordeling, zou ik het willen gebruiken.						
6. Als ik toegang had tot het overzichtssysteem voor beoordeling, denk ik dat ik het zou gebruiken.						
7. Ik ben van plan om het overzichtssysteem voor beoordeling te gebruiken in de toekomst.						

Gelieve de volgende vragen te beantwoorden:

	6-point Likert Scale					
	1 (Nooit)	2	3	4	5	6 (Altijd)
1. Bood het overzichtssysteem voor beoordeling de exacte informatie die u nodig hebt?						
2. Bood het overzichtssysteem voor beoordeling verslagen die precies lijken te zijn wat u nodig hebt?						
3. Bood het overzichtssysteem voor beoordeling voldoende informatie?						

Wat zijn volgens u de voordelen van het gebruiken van het overzichtssysteem voor beoordeling?

____

____

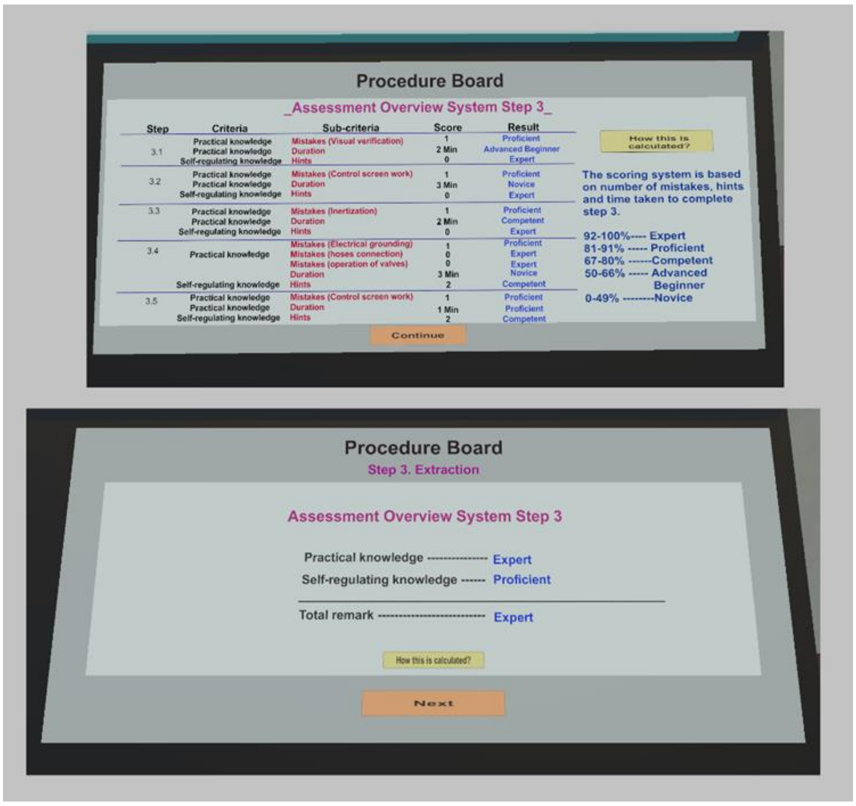
____

Wat zijn volgens u de nadelen van het gebruiken van het overzichtssysteem voor beoordeling?

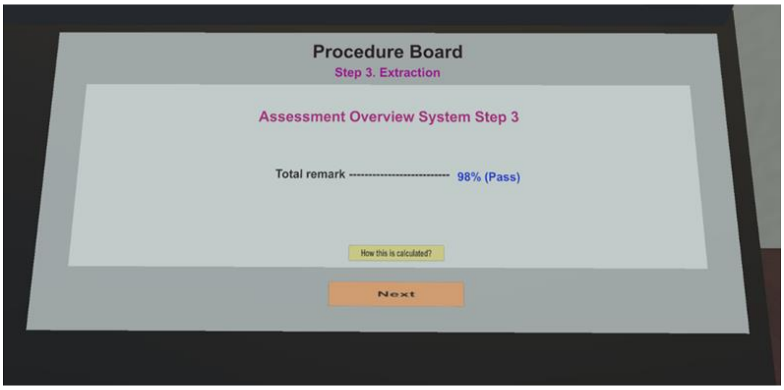
____

____

Hieronder vindt u de foto van het gedetailleerde overzichtssysteem voor beoordeling (1e foto) en het simpele overzichtssysteem voor beoordeling (2e foto)



gedetailleerde overzichtssysteem voor beoordeling (1e foto)



simpele overzichtssysteem voor beoordeling (2e foto)

Welk overzichtssysteem voor beoordeling heeft uw voorkeur in het VR-prototype? Gelieve een keuze te maken uit onderstaande opties.

- Ik heb liever de gedetailleerde versie (1e foto)
- Ik heb liever de simpele versie (2e foto)

Gelieve uit te leggen waarom dit overzichtssysteem voor beoordeling uw voorkeur heeft.
