A

Game-Based Assessment Framework for Engineering Education

by

Chioma Rita Udeozor

Student number: 19056434

Supervised by

Prof. Jarka Glassey and Dr Fernando Russo-Abegão

Funded by

European Union Horizon 2020 Research Framework Programme for Research and Innovation Number 812716

A Thesis Submitted to the School of Engineering

Newcastle University

For the Degree of a

Doctor of Philosophy

In

Chemical Engineering Education

March 2023

Abstract

Immersive technologies are gaining a lot of attention in various sectors for teaching and training. In the higher education sector, the interest in and the adoption of these technologies soared following the recent COVID-19 pandemic that limited face-to-face teaching. Engineering students were especially affected by these restrictions due to their inability to use the physical laboratories and workshops on campus. Engineers are trained to acquire both technical and non-technical skills necessary to solve current world issues. For the development of these skills, contextual learning environments are often required. Studies have shown that immersive technologies such as virtual reality (VR) and augmented reality (AR) are effective complements to current pedagogical tools due to their potential to enhance the development of complex skills and knowledge. One concern, however, with the use of these technologies for formal education is that educators are still treating them as “black boxes”. Educators lack the skills needed to design assessments for measuring learning in immersive environments. The research presented in this thesis aims to solve this problem by introducing an assessment framework that is underpinned by established pedagogical principles. This thesis also specifies best practices for the use of immersive technologies for higher education. To inform these outcomes, five empirical studies were carried out with engineering students and staff participants. Research designs such as mixed-methods, correlational and design-based research were adopted for these studies. The results provide useful insights into the views, behaviours and performance of engineering students in immersive environments. The results of the evaluation of the framework suggest that it is easy to use and useful for the intended purpose. Its application to the design of assessments for a VR application, an AR application and an educational digital game indicates that the framework is robust and could provide a structured basis for the design of assessments for immersive learning.

Dedication

I dedicate this work to my husband, Kelechi Udeozor, and our kids, Claire Udeozor and Bryan Udeozor. My heartfelt appreciation also goes to my parents and sister for their immeasurable support.

Acknowledgements

Looking back at my beginnings, the small town I grew up in, the people that I was surrounded by and the big dreams that I had, I can only now see what critical role my parents played in helping me get where I am today. I want to say a hearty thank you to my parents, Mr and Mrs Charles Onanike for all they do; their prayers, well wishes and support throughout this research project. The pride on their faces and in their words when they talk about my work give me reasons to fight harder each time.

I was able to put in my best work during this 3+ years of this research largely because of the support from my number one fan, my husband, Kelechi Udeozor. His sacrifices, listening ears, helping hands and encouragement were invaluable. I owe the success of this research to his candid support and interest in my work. I also want to acknowledge the tiny beings we have together. They have kept me on my toes and have made me stronger than I ever imagined I could be.

I would not have accomplished all these if not for one person who took a chance on me. I still remember how I danced with my then 2 years old daughter following an email from Prof Jarka Glassey. She gave me the opportunity to rediscover myself and dream bigger. She did not stop at that. She was always there, always available to answer questions, give feedback, and to offer solutions and guidance when needed. I have learned a lot more from her this past couple of years. I am grateful to you, Jarka. Thank you.

I also want to appreciate the whole team of the CHARMING project, particularly my co-supervisor, Fernando Russo-Abegao. He was always willing to roll up his sleeves and help out where necessary. He ensured that all our joint papers and this thesis were thoroughly reviewed and print-ready. Thank you, Fernando.

My appreciation also goes to my co-researchers who supported me in one way or the other. Through this project, I have come to meet extremely talented colleagues turned friends.

I am incredibly grateful to the European Research Commission that funded my research. The well-thought-through H2020 programme made it possible for me to visit and work with partners in different parts of Europe. Their generous compensation allowed me to focus on my research without any worries about my finances.

Last but most importantly, I am grateful to God for giving me health and life. There were many occasions when things would have gone wrong, but HE intervened. I am eternally thankful.

Table of Contents

List of Tables

List of Figures

Acronyms

.csv: comma-separated values

3D: three dimensional

AI: artificial intelligence

AR: augmented reality

ATT: average time on task

AVE: average variance extracted

BCa: bias-correlated and accelerated bootstrapping

BIs: behavioural intentions

CA: constructive alignment

CAVE: the cave automatic virtual experience

CB-SEM: covariance-based structural equation modelling

CHARMING: chemical engineering immersive learning

COVID-19: coronavirus disease

CR: composite reliability

DBR: design-based research

DEGs: digital entertainment games

DGBL: digital game-based learning

DGs: digital games

ECD: evidence-centred design

EE: energy expended

EIT-KIC: European Institute of Innovation & Technology – Knowledge and Innovation Community

EQF: European qualifications framework

ETS: educational testing services

EUH2020: European union horizon 2020

FCs: facilitating conditions

GBAF: game-based assessment framework

GBAF 2: revised game-based assessment framework

GPS: global positioning system

GRE: Graduate Record Examination

H: habit

H&S: health and safety

HCI: human computer interaction

HE: higher education

HM: hedonic motivation

HMDs: head-mounted devices

HP: health point

HTMT: Heterotrait-Monotrait ratio

IDs: identities

IEEE: Institute of Electrical and Electronics Engineers

ILO: intended learning outcome

IP: Internet Protocol

KU: Katholieke Universiteit

L&T: learning and teaching

LS: levels solved

MAR: mobile augmented reality

MCQs: multiple choice questions

MMORPG: massively multiple online role-playing game

N: sample size

n: sample size

N/A: not applicable

PBD: projection-based displays

PE: performance expectancy

PhD: Doctor of Philosophy

PLS-SEM: partial least square structural equation modelling

PPE: personal protective equipment

PV: price value

***P*-value**: statistical significance value

r: correlation coefficient

R^2: coefficient of determination

RQ: research question

s.d: standard deviation

SEM: structural equation modelling

SI: social influence

SOLO: structure of observed learning outcomes

TAM: Technology Acceptance Model

TOEFL: Test of English as a Foreign Language

UK: United Kingdom

USE: use behaviour

UTAUT: Unified Theory of Use and Acceptance of Technology

UTAUT2: extended Unified Theory of Use and Acceptance of Technology

VIF: variance inflation factor

VR: virtual reality

$\boldsymbol{\alpha}$: Cronbach’s alpha

$\boldsymbol{\beta}$**:** path coefficients

List of Publications and Conferences

1. Udeozor C., Russo Abegão F., & Glassey J.(2021). An Evaluation of the Relationship Between Perceptions and Performance of Students in a Serious Game. *Journal of Educational Computing Research*. doi:10.1177/07356331211036989
2. 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, DOI: 10.1080/23752696.2021.1951615
3. Udeozor, C., Russo-Abegão, F., & Glassey, J. (2021). Exploring Log Data for Behaviour and Solution Pattern Analyses in a Serious Game. In U. Bakan & S. Berkeley (Eds.), *Gamification and Social Networks in Education*. MacroWorld Pub. Ltd. https://doi.org/10.15340/978-625-00-0106-6_10
4. 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*. https://doi.org/10.1080/03043797.2022.2093168
5. Fornos, S., Udeozor, C., Glassey J., & Cermak, D. (2022). The CHEM Jam – How to Integrate a Game Creation Event in Curriculum-Based Engineering Education. *Education for Chemical Engineers*. https://doi.org/10.1016/j.ece.2022.04.001
6. Udeozor, C., Russo-Abegão, F., & Glassey, J. (2023). Perceptions and factors affecting the adoption of digital games for engineering education: a mixed-method research. *Int. Journal of Educational Technology in Higher Education.*
7. Domínguez Alfaro, J. L., Udeozor, C., Solmaz, S., Cermak-Sassenrath, D. (2022). Designing a Mobile Game for Introducing Learners to a Soap Making process. *16th European Conference on Games Based Learning, Lisbon*

8. Udeozor, C., Domínguez Alfaro J.L., & Glassey, J. (2023). Assessment framework for immersive learning: application and evaluation. *9th International Conference of the Immersive Learning Research Network.*

9. Udeozor, C., Russo-Abegão, F., & Glassey, J. (2023). Measuring Learning in Digital Games: Applying the Game-Based Assessment Framework. *British Journal of Educational Technology.*

10. Udeozor, C., Chan, P., Russo-Abegão, F., & Glassey, J. (2023). Assessment Framework for Virtual Reality, Augmented Reality and Digital Game-Based Learning. *International Journal of Educational Technology in Higher Education*

11. Solmaz, S., Cermak-Sassenrath, D., Domínguez Alfaro, J. L., & Udeozor, C. (2023). Bunno's Fabulous Soap-Making Challenge – Educational Concept, Game Design and Initial Evaluation. *Chemical Engineering and Technology*.
12. Jessica Lizeth Domínguez Alfaro & Chioma Udeozor (Submitted). Effectiveness of Augmented Reality Application for Science Education. *Education Sciences*.

Declaration: Publications 1, 2, 3, 4, 6, 8, 9, and 10 are products of the research presented in this thesis. Publications 5 and 7 were contributions made to the work of other authors who were part of the CHARMING project. All publications are in the field of immersive learning. The author of this thesis was responsible for planning and executing the research. They (the author this thesis) also collected, analysed and synthesised the data presented in the papers. They wrote, edited and submitted the manuscripts, too. For papers 2, 8 and 10, the author collaborated with colleagues on data collection. For paper 2, Toyoda collected and presented data from employees; Domínguez Alfaro developed the AR game and MCQs presented in paper 8, and Chan developed the VR game and MCQs presented in paper 10. Russo-Abegão and Glassey supervised the project and reviewed the manuscripts.

Conferences and Seminars

Poster presentations

Udeozor C, Russo Abegão F, Glassey J. (2020). Evaluation of the Use of a Serious Game for Chemical Engineering Education. *In: 14th European Conference in Games Based Learning (ECGBL 2020)*. Brighton, UK.

Udeozor C, Russo Abegão F, Glassey J. (2021). Relationship Between Perceptions and Experiences on the Performance of Students in a Serious Game. *In: 7th International Conference of the Immersive Learning Research Network.*

Extended abstract

Solmaz S, Cermak-Sassenrath D, Dominguez Alfaro JL, Arztmann M, Udeozor C. (2022). An Experimental Casual Game to Educate Players in Soap Production. *In: 13th European Congress of Chemical Engineering and 6th European Congress of Applied Biotechnology*.

Oral presentation

Udeozor C, Toyoda R., Russo Abegão F, Glassey J. (2020). Perceptions towards Virtual Reality Games for H&S Education and Training. *Newcastle University Learning and Teaching Conference*.

Udeozor C, Russo Abegão F, Glassey J. (2021). Behaviour and Solution Patterns in Chemical Engineering Education Game Log Data. *European Conference on Games-Based Learning***.**

Udeozor C, Russo Abegão F, Glassey J. (2021). Evaluation of Game-Based Learning for Engineering Education. *Newcastle University Postgraduate Research Conference.*

Udeozor C, Glassey J. (2021). Immersive Technologies for Chemical Engineering Education. *IChemE Education Special Interest Group. Chemical Engineering Education Online Symposium*.

Chapter 1. Introduction

The higher education (HE) sector is constantly evolving to meet the changing needs of industry employers, HE students and society. Globalisation, made possible by technological advances, is also putting pressure on institutions to constantly evolve to accommodate the changing demands of students whose backgrounds are increasingly distinct (McPhee & D'Esposito, 2018) due to differences in sociocultural status and academic preparation. As teaching and assessments are central to HE practices, efforts are constantly being made to restructure the curricula and ensure that students are satisfied with the experiences provided (Alpay, 2013). Digital technologies have often been embraced as tools that can be used to enhance the experiences of students, as well as support the workload of educators. Although HE institutions often promote the adoption of technologies across all academic disciplines, the medical and healthcare disciplines are almost always ahead of the others when it comes to adopting innovative digital learning and teaching (L&T) technologies (Paro *et al.*, 2022). The use of advanced digital technologies such as virtual reality in medicine can be traced as far back as the 1960s (Coiffet and Burdea, 2003). While this is not the case for most engineering education disciplines, engineers are always at the centre of these innovations, designing and bringing them to life.

Engineers are trained to sustainably utilise the limited resources available to improve quality of life. In the education of engineers, students are taught and supported to develop the competencies required to become professionals equipped to tackle the challenges of today and the future. Asides technical engineering competencies, four sets of core competencies (the four Cs) expected of professional engineers are critical thinking and problem-solving, collaboration and team building, effective communications, and creativity and innovation (Malheiro *et al.*, 2019). In addition to these competencies, the four Cs listed prior, Malheiro *et al*. (2019) included two other relevant competencies: socio-professional ethics, and sustainable development, as crucial to ensuring that future engineers contribute to the preservation of life on Earth and the overall well-being of people.

It is not always possible to teach these skills in traditional classroom settings with fixed physical spaces such as lecture theatres and the widely used lecture-based pedagogy. These traditional learning and teaching settings often pose challenges to replicating and updating real-world learning environments where students can develop and apply skills and knowledge to solving real-world problems. Contextual learning environments are often necessary to promote the development of the required technical and soft skills. To support students to develop these skills, educators rely on appropriate L&T methodologies that enhance active, situated and experiential learning which are, increasingly made possible by innovative digital technologies. Active learning can be considered as any instructional activity that requires students to perform meaningful activities (Prince, 2004). Situated learning theory presumes that learning is unintentional and happens within real world context, activity and culture (Lave & Wenger, 1991). It is the basis for experiential learning. Experiential learning emphasises the concrete experiences students gather from playing active roles in learning activities. Kolb and Kolb (2005) proposed an experiential learning framework that is widely adopted by educational practitioners. This framework purports that learning is a process (not an outcome) that requires the resolution of conflicts, results from interactions between the learner and their environments and is a process of knowledge creation that requires thinking, feeling, perceiving, behaving and relearning (Kolb & Kolb, 2005). Active, situated and experiential learning ground immersive learning and the work presented in this thesis. These L&T methodologies are very important in engineering education and are often achieved through problem-based learning and project-based learning (Duarte *et al.*, 2022; Mio *et al.*, 2019; McQuade *et al.*, 2020). Both learning approaches provide opportunities for students to develop the core professional competencies earlier mentioned while finding solutions to real-world problems (McQuade *et al.*, 2020). Innovative technologies are generally been relied upon to provide safe but realistic learning environments for engineering students to experiment and actively participate in learning activities (Amorim & Tavares de Azevedo, 2021; Duarte *et al.*, 2022).

Another learning theory that is promoted with immersive learning technologies and which grounds the work presented in this thesis is the constructivist learning theory. Constructivism, one of the widely adopted learning theories, is the view that students learn through active participation and reflection, and by updating their prior knowledge based on new information (Allen & Bickhard, 2022). Context, beliefs and attitudes of students are also said to affect learning (Bada, 2015). In addition to fostering constructivist, active and experiential learning, immersive applications such as DGs, VR and AR games and simulations, provide engaging learning environments that are intrinsically motivating and that enhance flow experience in the learning environment (Garris *et al.*, 2002; Prensky, 2003; Csikszentmihalyi, 1990). Flow is defined as the state of optimal experience that makes one engaged in a complex, goal-directed activity such that they lose the sense of time and self-consciousness (Shute & Rahimi, 2021). These attributes promote the engagement of students in gameplay activities while having them invest extensive amounts of time and energy into learning (Garris *et al.*, 2002). The enjoyment derived from the interactivity built into immersive learning environments makes them uniquely relevant for learning. With the ability to simulate real-world environments, engineering students can have access to industry-standard laboratories, process plants, and other environments that might be difficult to access physically in university settings. Immersive environments provide safe spaces for exploration and experimentation without putting the students, staff and property at risk. They are cheaper to use compared to building, maintaining and updating new physical learning environments or taking students on tours to industries. There are also some limitations to using these technologies, such as the time and technical skills required to develop good-quality applications.

As the global world moves towards the 4th Industrial Revolution with different technological trends reshaping industrial production, the demands on engineers are also changing. While the spread of automation of industrial processes and digitisation through the use of computers characterised the 3rd Industrial Revolution, the 4th Industrial Revolution is marked by digital transformation, artificial intelligence, industrial Internet of Things, cyber-security, simulation, extended reality and digitalisation. The implication for engineering education is that current educational practices may not be sufficient to prepare students for the demands of the current and future workplace. From the incorporation of games and simulations into current active learning strategies (Amorim & Tavares de Azevedo, 2021), to digital skills training (Keaveney *et al.*, 2021) and the use of learning factories (Erol *et al.*, 2016; Cooper *et al.*, 2020), several strategies have been suggested to better prepare students for the jobs ahead. In all, innovative digital technologies are expected to take centre stage in making these strategies possible. Games and simulations are increasingly employed to engage learners in realistic activities that enhance the acquisition of core professional engineering competencies (Udeozor, Toyoda, *et al.*, 2022). The ability of games to simulate interactive real-world environments makes them particularly useful for engineering education where access to real-world environments is limited for safety, cost and ethical reasons. Although simulations and games have been explored for engineering education in the past, technological advances and the challenges posed by the recent COVID-19 pandemic led to substantial interest in their use for engineering education.

Digital games (DGs) in classrooms can be traced as far back as the 1970s (Jones, 2017). Although initially met with mixed views due to the perceived negative effects on players, particularly regarding addictions and anti-social behaviours, the verdict has so far been in favour of the use of DGs for education (Klopfer *et al.*, 2009; Boyle *et al.*, 2011). In recent years, interest has shifted from DGs played on computers to more immersive gameplay experiences made possible with virtual and augmented reality technologies. Virtual reality (VR) and augmented reality (AR) were first seen in the 1960s as Morton Heiligi's Sensirama Simulator and Ivan Sutherland's Ultimate Display (Coiffet & Burdea, 2003; Sutherland, 1965). However, it was only a few years ago that the application of these technologies to education bloomed, owing to better computing power that enhanced visual rendering (Paro *et al.*, 2022). So far, VR and AR and most recently, mixed reality (MR) are being used to immerse learners in simulated virtual environments where they can interact in gameplay or other realistic learning activities. Detailed explanations and discussions about DGs, VR and AR will be given in Chapter 2.

The current boost in interest in immersive technologies, such as DGs, VR and AR, for education was also amplified by the COVID-19 pandemic that put restrictions on in-person L&T activities. Educators sought out possible alternatives to classroom and laboratory teaching activities that could engage students who had been forced to learn remotely. For engineering educators, there was a need to get students back into the laboratories, but the restrictions meant that alternative virtual laboratories had to be considered. Several institutions explored immersive technologies for this purpose (Glassey and Magalhães, 2020; Bhute *et al.*, 2022). Immersive technologies refer to those technologies that create a sense of immersion in a virtual world by blurring the line between the physical world and the digital or simulated world (Lee *et al.*, 2013). Immersive learning can be defined as the integration of educational content into immersive environments and their use for education and training. VR, MR and AR are considered synonymous with immersive technologies given that applications using these technologies enhance some level of immersion in digital worlds. Similar to AR, MR is used to describe a blend of the physical and virtual worlds. In MR, interactions between the physical and virtual worlds are possible whereas, in AR, virtual objects are simply superimposed onto physical environments (Park *et al.*, 2020). MR is not under consideration in this research due to its sparse use in education. For this thesis, the author argues that DGs played on devices other than VR and AR-enabled devices can also be considered immersive technologies albeit the level of immersion experienced is lower than in VR or AR environments. Therefore, DGs are used here to refer to games played on any electronic device other than VR or AR-enabled devices, for example, games played on computers, televisions, mobile devices or consoles. Distinctions between DGs and AR applications accessed via mobile phones, tablets, and personal and desktop computers will be made where necessary throughout this thesis. Immersive learning environments, the simulated or game-based environments offered by VR, AR and DG applications, provide active, situated and experiential learning environments that may promote active learning and reflection, important aspects of the constructivist learning theory (Shute *et al.*, 2017).

In terms of learning effectiveness, studies have reported similar or better outcomes with these technologies for engineering education (*e.g.* Criollo-C *et al*., 2021; Gómez-Tone *et al*., 2020; Lin and Wang, 2019; Suescún *et al*., 2018). In many of these cases, it was also important to understand user perceptions and views towards these technologies with mixed outcomes reported. Regardless of the positive outcomes expected and reported, immersive technologies have some limitations that could hinder adoption and effectiveness for education. For engineering education, designing new immersive learning applications that are relevant to the curriculum requires game design skills that very few educators have. Outsourcing these designs can also be expensive, and would require educators to invest time and resources to collaborate with game designers in order to develop pedagogically sound learning applications. AR and VR head mounted devices are also not always affordable at large scale, especially when these are needed for large classrooms. This can however be mitigated by group-based learning and by casting VR environment/interactions on large screens for others to see and participate in. Furthermore, these devices may not be suitable for all learners as they are known to cause motion sickness in some users (Checa & Bustillo, 2020). Regardless, with technological advances, it is expected that most of these challenges will be minimised in the near future. Devices are expected to become more affordable, visual rendering will be improved to reduce motion sickness, and authoring tools for game designs should become available to enable educators to create immersive learning applications with little or no professional game design skills (Ardiny & Khanmirza, 2018).

A growing number of studies on the pedagogical implications of immersive technologies for L&T exist as a result of interest in their use for formal education. Design considerations for immersive learning environments attracted more attention with a lot more research interest and output (*e.g.* Garcia Fracaro *et al*., 2021; Solmaz, Dominguez Alfaro, Santos, Van Puyvelde, and Van Gerven, 2021; Solmaz and Van Gerven, 2021). Many recommendations were published on the best ways to design learning content in DGs, VR and AR applications for maximum impact, particularly as outcomes from the CHARMING project, of which this research is a part. Others focused on strategies for incorporating immersive technologies into existing classroom instructional activities (Wouters & van Oostendorp, 2017). Fewer studies discussed assessment implications when teaching with these technologies, with most calling for research into assessments for immersive learning

(Connolly, Stansfield and Hainey, 2009; Bellotti *et al*., 2013; Kumar *et al*., 2021). Studies showed that although educators are generally receptive to the use of immersive technologies to improve learning (Noraddin & Kian, 2014; Stieler-Hunt & Jones, 2015; Khukalenko *et al.*, 2022; Beavis *et al.*, 2014; Razak *et al.*, 2012), concerns over the assessments of learning (Razak *et al.*, 2012), and the need for adequate classrooms pedagogy-relevant training (Khukalenko *et al.*, 2022), ranked highest among the challenges outlined by educators.

Immersive technologies provide unique opportunities to measure what students *can do* by applying their conceptual understanding of subject matters, rather than what *they say* or their responses to traditional exam questions. Using immersive technologies, students can demonstrate higher-order cognitive processes in their interactions within the environments and these can be used to measure complex knowledge and skills that would otherwise be difficult to measure. However, because immersive technologies are relatively new in the classrooms, they are still being treated as “black boxes” by educators and researchers (de Klerk *et al.*, 2015). Previous reviews of the literature showed that traditional assessment methods such as the use of tests are still overwhelmingly used with these technologies (Li *et al.*, 2017; Udeozor, Toyoda, *et al.*, 2022; Kittur & Islam, 2021). There is ongoing criticism of the use of exams for measuring learning in HE because of the challenges of designing authentic assessment items that measure higher-order cognitive processes (Villarroel *et al.*, 2019). While immersive learning environments provide a viable alternative to exams, the process of designing and implementing assessments in these environments is not yet clear to educators (Kumar *et al.*, 2021; Connolly *et al.*, 2009; Bellotti *et al.*, 2013).

In the health and medical science domain where immersive learning has existed for some time now, subjective and objective assessment methods have been generally used to determine the competencies of trainees. Pre- and post-intervention tests, biometrics such as eye tracking and other automatically captured metrics of learner interactions are often used to objectively measure surgical skills performance (Topalli & Cagiltay, 2019; Xie *et al.*, 2021; Menekse Dalveren & Cagiltay, 2020). Subjective methods such as the use of self-reports from trainees and observations by experts are also considered invaluable and used to determine the effectiveness of immersive training (Xie *et al*., 2021). With subjective and conventional assessment methods still prevalently used in the medical science domain for skills assessment, it is evident that although progress is being made in the use of immersive technologies for education, more needs to be done in terms of the measurement of learning in immersive environments (Menekse Dalveren & Cagiltay, 2020). Given where the medical science domain is in the move towards immersive learning and assessment, the engineering education domain is not too far off. However, with the limitations of the current assessment methodologies used in most domains, the engineering education domain might benefit from identifying or developing appropriate and more reliable assessment alternatives as is currently being done in the medical domain (Menekse Dalveren & Cagiltay, 2020). There are no known frameworks developed specifically for assessment designs for immersive medical and surgical skills training. As with other educational domains, existing assessment frameworks not specific to immersive learning are adopted. In the medical domain, frameworks such as the objective structured assessment of technical skill (OSATS) and the JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS) are being applied to VR and simulations-based training (Zia & Essa, 2018; Martin *et al.*, 1997). These are specific to medical skills assessment, hence would not be appropriate for assessments in other domains.

In other educational domains, including engineering, a few assessment design frameworks and guidelines such as the Evidence-Centred Design (ECD) and Information Trails have been identified and applied to assessments in educational DGs. However, they are generally considered complex to use, effective for use during the design phase of the environment and require advanced statistical skills (Kim, Almond and Shute, 2016; Westera, 2019; Westera *et al*., 2020). These conditions make it challenging for educators to use these frameworks. The inability to accurately measure intended outcomes and make claims about what students “know” and “can do” could potentially limit wider acceptance and adoption of immersive learning technologies by educators. To begin to address the assessment challenges associated with using immersive technologies for L&T, an assessment design framework relevant to immersive learning environments is required. An educator-friendly framework that demonstrates assessment design considerations and procedures relevant to immersive learning environments, which is the focus of this thesis, is needed. This work presents an assessment framework based on the Constructive Alignment principles and the ECD framework. Details of assessment implications for immersive learning and assessment considerations made for the work presented herein are discussed in detail in sections 2.5 and 6.4, respectively.

1.1 Research Problems

Three major research problems relating to assessments in immersive environments addressed by this thesis are:

1. The interest in the use of immersive technologies for teaching and assessments in HE is growing with a growing number of research studies into pedagogical implications.

However, it remains unclear what the views of engineering students are on this subject and how these could influence assessments.

2. Immersive learning environments are different from most conventional learning activities. How engineering students behave in these novel learning environments and the implications of their behaviour on the assessments, are still unknown.
3. A lack of knowledge about the potential methods of assessing performance when using immersive learning applications exists. There is also a lack of educator-friendly guidelines for designing and implementing assessments when teaching with immersive technologies.

1.2 Proposed Approach and Thesis Structure

To address the above-outlined research problems, a literature review of current educational uses of DGs, VR and AR in the engineering domain is conducted. The perceptions of engineering students towards immersive technologies for education are then explored, which provides answers to the first research problem. A follow-up study is carried out to further investigate how the perceptions and experiences of students affect their performance in immersive learning environments. An exploration of the gameplay behaviours of engineering students in an educational DG is carried out in response to the second research problem. Furthermore, an in-game assessment method for measuring the performance of students in immersive learning environments is explored. The thesis concludes with the development of a game-based assessment framework useful for designing assessments for immersive learning. This thesis also describes the evaluation and implementation of the framework for the design of assessments for a DG, a VR and an AR application. The framework highlights relevant components required for assessments in immersive environments and their interrelationships. Stepwise applications of this framework to the design of assessments for immersive learning environments are detailed to illustrate its practicality.

This thesis is therefore divided into eight chapters as depicted in figure 1.1. The focus of each chapter is as follows:

Chapter 2: This chapter presents a systematic review of literature on the reported uses of DGs, VR and AR in engineering education disciplines within the past ten years. It describes the DGs used as well as the AR and VR devices adopted. It also discusses the experimental research designs implemented, the assessment methods used, and the outcomes reported by the studies.

Chapter 3: This chapter evaluates the perceptions of engineering students towards the use of DGs for education. Adopting a mixed-methods research design with surveys and focus-group discussions, key factors that would influence the adoption of DGs by engineering students are identified. The pedagogical implications of the results of this study are also described.

Chapter 4: To follow up on the concerns raised by students regarding gaming experience interference on performance in educational DGs, a correlational study is carried out. This chapter uses a correlational research methodology to identify relationships between the gaming experiences of engineering students, their perceptions of the use of DGs for education and their performance in an educational DG. Pedagogical implications of the finding of this chapter are discussed.

Chapter 5: This chapter explores the gameplay behaviour of engineering students in an educational DG. It uses a grounded theory research design to present an understanding of how these groups of students behave in immersive learning environments and the implications their behaviours could have on assessment.

Chapter 6: This chapter details the development of a game-based assessment framework (GBAF). The GBAF is developed primarily to guide educators through the process of designing and implementing assessments when teaching with immersive technologies. An initial evaluation of the framework is reported as well as the need for improvements.

Chapter 7: This chapter presents a revised framework, the GBAF 2, which takes into account the limitations of the original framework. It also presents an evaluation of the framework and its application to the design of assessments for a DG, VR and AR game. Best practices for assessment implementations are provided.

Chapter 8: This chapter reflects on the research presented, highlighting its contributions, strengths and limitations. It also presents propositions for future research.

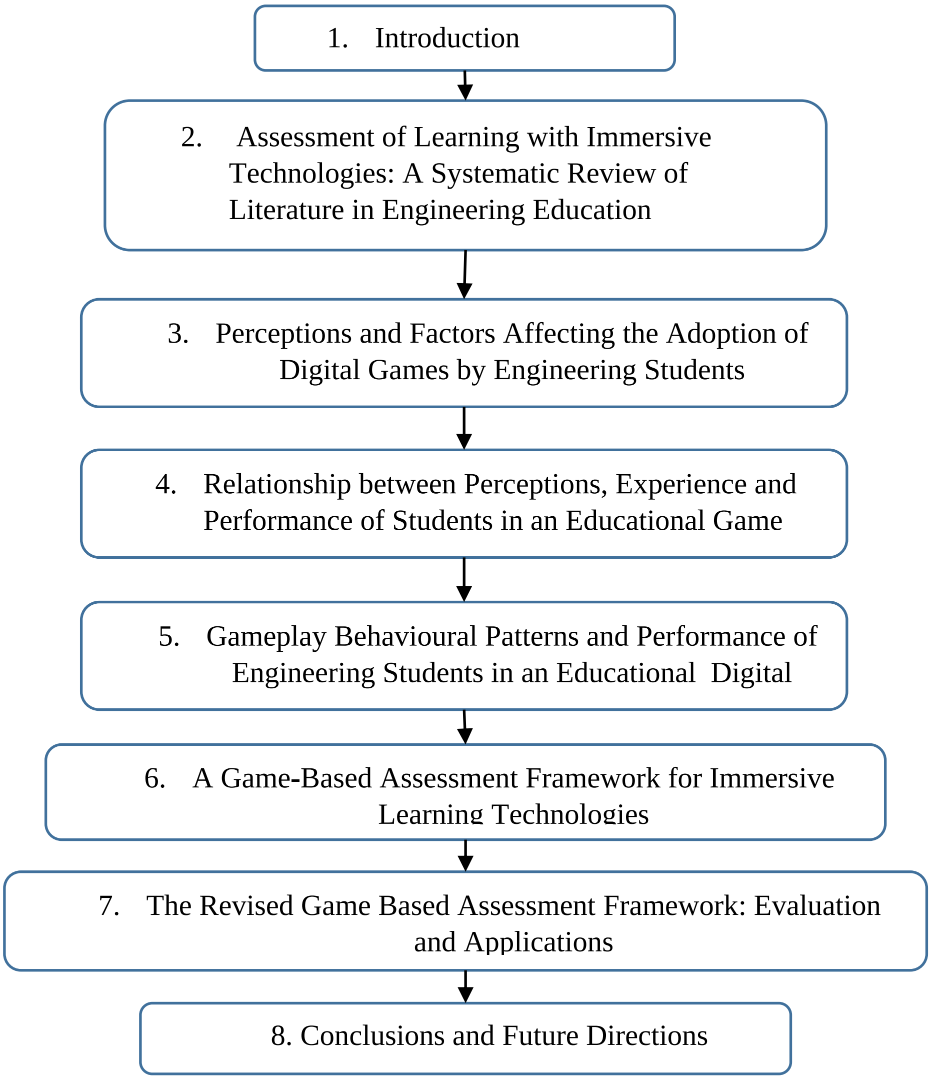


Figure 1.1: Outline of thesis structure

Chapter 2. Assessments of Learning in Immersive Environments: A Systematic Review of Literature in Engineering Education

The previous chapter provided an overview of the applications and benefits of immersive technologies for education. A background of the theories, concepts and terms associated with immersive technologies is first provided in this chapter. Subsequently, the current uses of these technologies for engineering education are explored. Previous reviews have focused on identifying the applications of serious games (Kittur & Islam, 2021), virtual reality (VR) (Vásquez-Carbonell, 2022b) and augmented reality (AR) (Vásquez-Carbonell, 2022a) in engineering education, separately. The serious games review provided non-exhaustive data as it was conducted on publications on digital games (DGs) and non-digital games from one database. The VR and AR reviews focused on uncovering trends such as countries leading the adoption of immersive technologies, funding types, hardware and software used. There was little to no emphasis on learning and assessment considerations reported in the reviewed papers. The current review provides a holistic overview of the uses of all three technologies for learning with a particular interest in assessment designs and methods used in engineering education, factors that were not taken into adequate consideration in previous reviews. It also follows a systematic review process, searching five databases for papers published within the last decade on the use of DGs, VR and AR applications for engineering education. From almost 6000 papers returned by the searches, only the 66 that met the inclusion criteria (*cf.* section 2.6) were analysed in this review. The findings show the application of immersive technologies to a wide range of engineering disciplines. Different assessment methods were used in the reviewed papers to measure the learning effectiveness of immersive technologies for engineering education. External assessment methods, specifically tests, were identified as being used by a large proportion of the papers, with only a few papers reporting the use of embedded assessment methods. The outcomes of the assessments were generally positive, but with different assessment methods used, and the limitations posed by the sample sizes and the experimental designs adopted, it would not be possible to generalise these outcomes. The findings of this chapter have implications for the wider use and adoption of immersive technologies for learning and assessment in higher education (HE).

2.1 Introduction

Integral to everyday classroom activities in HE are the components of teaching and assessments. From teacher-led lecturing to workshops, tutorials, laboratory sessions, simulations and other technology-aided active learning approaches, learning and teaching (L&T) in HE takes many forms. Lecturing remains the most popular teaching method in HE (Roberts, 2019). Nevertheless, technological advances, fierce competition among institutions, as well as the need to meet the growing demands of students and employers, are fuelling the adoption of innovative and immersive learning technologies such as DGs, VR and AR (Bahja *et al*., 2021). Regardless of the L&T method used and the technology adopted, to understand if and how much students actually learned, educators often rely on assessments (Fry *et al.*, 2003). The goal of this chapter is therefore, to first outline some background information on educational assessments, introduce the immersive learning environments under study, and then present the state-of-the-art uses of these technologies for learning and assessment in engineering education.

2.2 Assessment

The term "assessment" has been defined by different authors to mean different things. Two interesting definitions came from the works of Mislevy, Steinberg, and Almond (2003) and Reeves and Hedberg (2009). Mislevy, Steinberg, and Almond (2003, p. 4) define assessment as "*a machine for reasoning about what students know, can do, or have accomplished, based on a handful of things they say, do, or make in particular settings*". Assessment is also referred to as "*activities focused on measuring characteristics of human learners (their learning, motivation, attitudes, etc*.)" (Reeves and Hedberg, 2009, p. 235). It is necessary to point out that in both definitions, references are made to the measurement of the competencies of learners. It is not uncommon to find the terms "assessment" and "evaluation" interchangeably used in educational literature. These terms are arguably different. Evaluations in HE are generally associated with estimating the value of curricula, projects or programmes, whereas educational or learning assessments are used to measure the learning achievements of students (Peeters & Schmude, 2020; Reeves & Hedberg, 2009). In this thesis, therefore, this distinction between assessment and evaluation is adopted. Assessment is hence used here to refer to the measurement of the learning or performance of students. Evaluation, on the other hand, is used here to describe the measurement of the performance of non-person entities such as interventions, products or programmes.

2.2.1 Formative and summative assessments

Assessments can be classified as either formative or summative. The difference between these two assessment types relates to how the results or data from the assessment are used. Summative assessments are administered to students periodically, such as at the end of an instruction unit or course, to determine their proficiencies at a given time. The results from summative assessments are used to ascertain proficiencies in learning outcomes and to judge performances (Dixson & Worrell, 2016). Summative assessments are generally high-stake assessments that are almost always graded and less frequently administered (Dixson & Worrell, 2016).

Although immersive technologies are being used for summative assessments, particularly for graduate recruitments, these tools are believed to have the greatest potential to support formative assessments (Shute *et al.*, 2018). Formative assessments are specifically used to collect relevant data from students for the purpose of improving learning (Dixson & Worrell, 2016). The results from formative assessments are used to give feedback to students, address misconceptions and modify L&T activities in time to improve learning. Formative assessments are sometimes graded, but these are typically expressed in words as feedback rather than marks and are often used for diagnostic purposes (Fry *et al.*, 2003). While formative assessments are generally used to improve learning, summative assessments are often used to judge proficiency.

Feedback is essential for learning because it provides students with the necessary information about their performance and how to improve in the future. Feedback is defined as "*a dialogic process in which learners make sense of information from varied sources and use it to enhance the quality of their work or learning strategies.*" (Carless, 2015, p.192). Feedback can be used for different purposes, however, to be effective, feedback has to be goal-oriented, actionable, tangible and transparent, personalised, consistent, ongoing, and timely (Wiggins, 2012). A huge benefit of immersive technologies for learning is their ability to consistently provide immediate feedback that is tailored to individual needs.

2.2.2 Validity, reliability and fairness in assessment

Whether intended for formative or summative assessments of learning, an assessment has to be valid, reliable and fair, to be effective. Validity describes the degree to which theoretical rationales and evidence support the appropriateness and adequacy of the inferences made based on test scores (Messick, 1987). An assessment is deemed valid when items of the assessment are appropriate for the intended purpose, that is to say when assessment tasks measure what is intended. Three aspects of validity commonly taken into consideration during assessment designs are face or content, construct, and impact validity (Fry *et al.*, 2003). Fry and colleagues describe face validity as the adequacy of the content of an assessment for the intended audience and their level. Construct validity is said to do with the nature of the broader constructs tested, while impact validity concerns the effect an assessment procedure has on the behaviour of learners (Fry *et al.*, 2003).

Whereas validity is judged qualitatively, the reliability of an assessment is quantitatively measured (Fry *et al.*, 2003). The length of the assessment items, the examiners, and individual assessment items could affect the reliability of an assessment. Reliability, therefore, is the degree to which an assessment is dependable, that is, the degree to which an assessment given to a particular student consistently produces similar assessment scores regardless of when the assessment was taken or scored (Moskal, Leydens and Pavelich, 2002; Driessen *et al.*, 2005). Some of the ways to assess reliability include internal consistency measurements using Cronbach's alpha or split-half tests, temporal stability measurements through test-retest, and interrater reliability using percentage agreement, Phi scores, and Kappa scores (Cook & Beckman, 2006).

In addition to being valid and reliable, assessments have to be fair to all individuals and different groups of individuals. Fairness can be achieved by using different assessment methods or different marking tendencies (Fry *et al.*, 2003). Fairness in access to schooling and curriculum opportunities is thought to provide a level playing field for all students, which in turn allows for genuinely fair assessments (Wyatt-Smith & Cumming, 2009). Fairness is considered interchangeable with equity. Wyatt-Smith and Cumming (2009) argue that fairness should be viewed as a sociocultural issue rather than a technical one. They added that fairness should not be treated as a separate concept but should be embedded within validity arguments that judge assessments based on constructs assessed, and the inferences/actions taken based on the results.

2.2.3 Learning outcomes

In HE, L&T processes typically follow the principle of Constructive Alignment that demands explicit connections between the aim of the learning activities (expressed as learning objectives), the L&T activities and assessment tasks designed to measure achievements on the set learning outcomes (Biggs & Tang, 2010). Fundamental to the design of valid assessments is the identification and articulation of the intended learning outcomes (ILOs).

According to the revised European Qualifications Framework (EQF), learning outcomes are statements of what a student knows, understands and is able to do following a learning activity (Bartolo, 2017). Learning outcomes are often defined in terms of knowledge, skills and competence. The EQF defines *knowledge* as the outcome of the absorption of information through learning (Bartolo, 2017). It is the body of facts, theories, principles and practices associated with an aspect of work or study. *Skills* are defined as the ability to apply and use knowledge to accomplish tasks and solve problems. Skills can be cognitive (requiring creative and logical thinking), or practical (requiring manual dexterity and use of materials, methods and instruments) in the EQF context. Lastly, *competence* is the demonstrable ability to use knowledge, skills and social, personal and/or methodological capabilities in study or work instances (Bartolo, 2017).

To appropriately define measurable learning outcomes for educational purposes, taxonomies are considered relevant (Biggs and Tang, 2010). There are two commonly used taxonomies for instructional objectives that are useful for drafting learning outcomes: the structure of observed learning outcomes (SOLO) taxonomy (Biggs & Collis, 1982) and Bloom's taxonomy by Benjamin Bloom (Anderson, 2013). Both taxonomies provide frameworks for instructional designs as well as guides for assessment designs. The SOLO taxonomy offers a means of classifying learning outcomes in terms of their complexity and quality of output. Bloom's taxonomy, however, is well known and used in educational settings as a way of classifying learning outcomes by the complexity of the cognitive process required. Unlike the SOLO taxonomy, Bloom's taxonomy is believed to capture nearly all possible cognitive educational objectives making it very useful for educators (Mosel*ey et al.*, 2005). According to the revised Bloom's taxonomy, learning outcomes include both knowledge and cognitive process dimensions. On the knowledge dimension, four types of knowledge are presented. These are factual, conceptual, procedural and metacognitive (Anderson, 2013) as shown in table 2.1. In engineering education, immersive technologies have been used by students to gain factual knowledge of sustainable building design practices (Dib & Adamo-Villani, 2014), conceptual knowledge of electrical currents (Urbano *et al.*, 2020) and procedural knowledge of life cycle assessment in manufacturing (Perini, Oliveira, *et al.*, 2018).

The cognitive process dimension, on the other hand, consists of six categories commonly found on lists of ILOs. These include remembering, understanding, applying, analysing, evaluating and creating, in order of complexity. To define valid and measurable learning outcomes, verbs associated with each category must be identified and used appropriately (Anderson, 2013). Central to the argument for immersive learning, is the understanding that these technologies enable the learning and assessment of higher-order cognitive processes (Voorhis & Paris, 2019; Rice, 2007). Table 2.2 outlines all six categories of cognitive process dimensions and the associated verbs.

Table 2.1: Knowledge dimension: types and subtypes (adapted from Anderson (2013)).

Knowledge Type	**Definition**	**Subtypes**
Factual knowledge	Knowledge of discrete, isolated pieces of information.	Knowledge of terminology Knowledge of specific details.
Conceptual Knowledge	Knowledge of more complex, organised knowledge forms.	Knowledge of classification and categories. Knowledge of theories, models and structures. Knowledge of principles and generalisation.
Procedural Knowledge	Knowledge of how to perform a task.	Knowledge of subject-specific techniques and methods. Knowledge of criteria for determining when to use suitable procedures. Knowledge of subject-specific techniques and methods.
Metacognitive Knowledge	Knowledge and awareness about self-cognition and about cognition in general.	Strategic knowledge Self-knowledge Knowledge about cognitive tasks, including appropriate contextual and conditional knowledge.

2.2.4 Authentic assessment

Typically, assessments in HE are administered as closed-book/open-book exams in the forms of multiple-choice tests, short answer questions, long essays, practicals or laboratory reports, and portfolios, designed to measure specific learning outcomes. Over the years, these assessment methods have transitioned from traditional paper-based tests to technology-based educational assessments. Bennett (2015) conceptualised the evolution of three generations of technology-based assessments. The first generation of assessments, based mainly on information technology infrastructure building activities, follows a transition from paper-based development and delivery of assessment to the use of computers. Assessment here is said to closely resemble traditional tests but may be delivered via computers and laptops. In the second generation, the exploration of the applications of technology to assessments is often the case. Generally, there is an emphasis on using non-traditional item formats such as multimedia content, essays and short response items and increasing attempts to measure new constructs. The third generation of assessment is one of reinventions. Assessment here is driven by substance and purpose, rather than by technology, as in the previous generations. Complex simulations like VR and AR simulations and other performance-based tasks that simulate real-world environments are often used.

Table 2.2: Cognitive process dimension: categories and associated verbs (adopted from Anderson, 2013).

Cognitive Process Category	**Definition**	**Associated Verbs**
Remembering	Retrieve relevant knowledge from long-term memory.	Recognising Recalling
Understanding	Construct meaning from instructional messages, including oral, written, graphic communication.	Interpreting Exemplifying Classifying Summarising Inferring Comparing Explaining
Applying	Carry out or use procedures in a given situation.	Executing Implementing
Analysing	Break material into constituent parts and determine how parts relate to one another and to an overall structure and purpose.	Differentiating Organising Attributing
Evaluating	Make judgement based on criteria and standards.	Checking Critiquing
Creating	Put elements together to form a coherent or functional whole; reorganise elements into new patterns.	Generating Planning Producing

The growing concerns about the validity of the use of some of the traditional assessment methods such as multiple-choice questions to measure higher-order cognitive processes in HE have led to calls for more authentic performance-based assessment methods (Villarroel *et al.*, 2019). Assessments are thought to be authentic when they are based on the activities of students that replicate real-world work or tasks (Svinicki, 2004; McArthur, 2022). According to Ashford-Rowe, Herrington, and Brown (2014), authentic assessments should be challenging, performance or product-outcome based, must ensure knowledge transfer, enhance self-reflection and self-assessment, be contextual, and accurate, and should encourage discussions and collaboration. These elements considered critical for authentic assessments are not always easy to integrate within traditional classroom assessments but can be achieved through problem-and project-based learning and assessments (Merrett, 2022). Authenticity is believed to be a continuum with most traditional assessment methods falling on one end of the continuum and simulated or real-world assessments found on the other end (Svinicki, 2004). While traditional assessments often require students to provide one correct answer, authentic assessments are aimed at measuring the quality of responses and justifications made by students (Svinicki, 2004). In engineering education, authentic assessment could be achieved through group work, design projects, industry-based projects or other performance-based tasks that provide students with real-world problems where relevant skills and knowledge can be learned, applied and assessed.

Authentic assessments have a significant number of benefits as well as drawbacks compared to formal classroom assessments. On the plus side, they can enhance the motivation of students when designed with the skills and interests of the students in mind. Face and construct validity, and transferability of skills are some of the other strengths of authentic assessments (Svinicki, 2004). As these assessments mirror real-world tasks, it is often easy for both students and evaluators to identify the links between the assessment tasks and the measured competence or construct. In HE, designing assessments that mirror what students would do post-graduation would likely be met with more enthusiasm compared to designing less authentic assessments. Nonetheless, the time and effort required to design and implement authentic assessments is a major drawback to their use. Designing tasks for authentic assessments is nontrivial, and often requires educators to create and validate a variety of assessment items for each learning outcome assessed. Another drawback of authentic assessments is the grading of these assessments. In addition to the time and effort required for grading, the reliability of the grading process also possesses some challenges. Given that there is often no single correct answer to the tests, ensuring consistency in the grading of the works of students is not always easy. However, with their use in clinical practices and their growing adoption in other HE domains, much more is being known about improving the quality of these sorts of assessments (Menekse Dalveren and Cagiltay, 2020; Xi*e et al.*, 2021). Comparing the performances of students is also challenging with authentic assessment as no two work products are the same (Svinicki, 2004). This could have implications for the use of authentic assessments for high-stake exams where the performances of students are sometimes compared (Dixson & Worrell, 2016). Besides the challenges authentic assessments pose to educators, they can be strenuous for students too. With the demanding workload of HE students, having to complete authentic assessments in all their modules will significantly increase their workload and could result in higher levels of dissatisfaction from students. Notwithstanding these challenges, authentic assessments have the potential to enhance the learning experiences of students as well as their employability post-graduation (Svinicki, 2004; Villarroel *et al.*, 2019). Immersive technologies can potentially facilitate authentic assessment designs and delivery in HE.

2.3 Immersive Learning Technologies

2.3.1 Digital games (DGs)

Game-based learning pedagogy, unlike most traditional pedagogies, is believed to foster active participation and engagement of students in learning activities. Games are described as structured or organised plays that are enjoyable and have defined goals (Klopfer *et al.*, 2009). Other elements of games could include fantasy, rules, challenges, feedback, competition, cooperation, control, and storyline (Prensky, 2003; Garris *et al.*, 2002). Although these elements characterise a game, they are not always present in every game. Gameplay construct lies in the interplay between game narratives and game mechanics. Game narratives can be described as the backdrop plot, mission, or storyline that provides an immersive experience to players during interactions in the game environments (Ke, 2016). Not all games tell a story but all games are made up of a set of structured activities comprised of rules and actions known as game mechanics (Ke, 2016).

Game mechanics and narrative are layers that make up a game and are often used to classify games into genres. As in literature, media, and arts, attempts have been made to classify games into genres that would allow players to identify games of interest. Some of the popular classifications of games were made by Wolf (2001) and Apperley (2006). Wolf (2001) presented 42 classes of game genres based on interactivity and iconography. These include board games, adventure, fighting, capturing, platform, puzzle, role-playing and simulations. Apperley (2006), on the other hand, presented four genres that can be used to classify games based on their inherent similarities and interactivity within each game. These game genres include the simulation genre consisting of games that are authentic rather than simply entertaining, the strategy genre for games that require constant attention and performance of players, the action genre for hyper-performative games, and lastly, the role-playing genre which is closely linked to the literary genre of fantasy (Apperley, 2006). These classifications of genres by Wolf and Apperley were targeted at video games, however, a later but non-exhaustive list of educational game genres was presented in Ke's (2016) systematic review paper. These include casual puzzles, action, adventure, strategy, role-playing, simulation and construction game genres. Simulation games are reportedly the most popular for education and training, as games in this genre are designed to simulate real-world situations while incorporating game elements like competition and cooperation (Connolly *et al.*, 2012; Boyle *et al.*, 2016). Simulation games are also the most popular game genre in engineering education and are considered to have promising applications in the domain (Deshpande & Huang, 2011; Kittur & Islam, 2021).

Games used in classrooms include non-DGs, such as traditional board games, and DGs which involve the use of electronic devices like mobile phones, televisions, computers, and most recently augmented and virtual reality-enabled devices. Advances in technology and the relatively low cost of devices have led the drive toward DGs to provide highly interactive learning environments (Slimani *et al.*, 2018). Although DGs are any games that can be played on electronic devices, for the purpose of this thesis, DGs are used to refer to games played on mobile phones, computers, consoles or televisions, as shown in figure 2.1. Digital game-based learning (DGBL), an instructional method that integrates educational content into DGs (Bahadoorsingh *et al.*, 2016), is reportedly used in diverse educational domains (Boyle *et al.*, 2016; Connolly *et al.*, 2012). Often, gamification is considered synonymous with game-based learning. However, this is not necessarily the case as gamification involves the application of game elements to non-game classroom activities to make them more engaging (Plass *et al.*, 2015). Such game elements include points, trophies or badges given to students for work done or progress made in the everyday classroom learning activity. Game-based learning, on the other hand, involves the use of educationally relevant games for classroom learning. Both game-based learning and gamification are used in engineering education, but there is an increasing preference for DGs (Bodnar *et al.*, 2016, 2021).

DGs used for learning can be categorised into two groups: entertainment games and educational or serious games. Digital entertainment games (DEGs) are DGs designed specifically for fun, recreational or entertainment purposes. Educational games and serious games, on the other hand, are games primarily designed with specific learning outcomes in mind other than just fun. These games are often designed to promote learning and behavioural changes (Connolly *et al.*, 2012). Both DEGs and educational games are effective learning environments, however, there are some advantages and limitations to using either for education. Since DEGs are usually developed by gaming companies with larger budgets, these games are often high-end products that are very engaging and readily available. On the flip side, DEGs can be expensive, have complex interfaces, and steep learning curves, and often do not match required curriculum outcomes (Whitton, 2009). In contrast, educational games are designed for learning, with a focus on intended learning outcomes at the design phase and often have good links to the curriculum. Educational games created specifically for learning may nonetheless fail to engage students (Squire, 2003), and creating these games can be expensive and time demanding. As a compromise, many DEGs now come with creation engines that make modifications for educational use possible (de Freitas, 2006). Studies have successfully used entertainment games (Bahadoorsingh *et al.*, 2016), educational games (Smith and Chan, 2017) as well as modified DEGs (Coller & Scott, 2009) for teaching and training in different engineering disciplines.

Figure 2.1: Screenshot of digital games played on a computer, mobile phone, console and television.

2.3.2 Virtual reality (VR)

Due to advances in technology, players of DGs can now be fully immersed in virtual worlds that offer more realistic experiences. For educational and training purposes, immersive VR and AR applications offer additional benefits over conventional 2D applications given their abilities to immerse learners in realistic virtual environments. VR is defined by Coiffet and Burdea (2003) as " a *high-end user-computer interface that involves real-time simulation and interactions through multiple sensorial channel. These sensorial modalities are visual, auditory, tactile, smell, and taste*". Virtual world, immersion, interactivity, and sensory feedback, referred to as the key elements of VR experience, encompass its full meaning (Sherman & Craig, 2003). VR brings to "life" an imaginary world that may or may not exist in reality, which can be experienced by users. It immerses one in a virtual environment, bringing about a sense of presence, physically and mentally, in a simulated virtual environment. Interactivity and sensory feedback elevate the sense of immersion in VR, resulting in "difficult-to-forget" experiences that make it particularly useful for training and education (Chittaro & Buttussi, 2015). Immersion in VR can be viewed as a continuum from non-immersive or desktop VR that displays virtual environments on personal computers (PCs), to semi-immersive VR on projection-based displays (PBD) such as PowerWalls and the cave automatic virtual experience (CAVE) system, to fully immersive head-mounted devices (HMDs) enabled VR (Cronin, 1997; Feng *et al.*, 2018). Both semi- and fully immersive VR offer a higher sense of presence in the virtual environment compared to desktop VR. Therefore, for the sake of this thesis, semi-immersive VR and fully immersive VR are considered immersive VR.

VR is not entirely a new concept. It dates back to the 1960s to Morton Heilig's invention of Sensorama Simulator (Coiffet & Burdea, 2003) and Ivan Sutherland's concept of Ultimate Display (Sutherland, 1965). In the past, the use of VR has been restricted to the military and for research due to high costs (Youngblut, 1998). Recently, the use of immersive technologies can be seen in every sector of the economy, including education. Technological advances, affordable costs of devices and the recent COVID-19 pandemic are fuelling interest in research and the application of VR for education. Desktop computers are comparatively cheaper than CAVE systems and HMDs. CAVE systems are said to offer higher resolutions per square degree of a visual angle compared to HMDs, however, these take up large amounts of space, require expensive high-resolution projectors and offer limited interactivity (Havig *et al.*, 2011). HMDs, on the other hand, take up less physical space, are not as expensive as the CAVE system and are not limited by physical space as with the CAVE system. One major downside of HMDs is that they can cause motion sickness (Havig *et al.*, 2011; Hettinger & Riccio, 1992), but this limitation is expected to diminish over time with advances in technology. Both HMDs and CAVE systems have been found to be effective for engineering education (Alhalabi, 2016). Figure 2.2 shows pictures of immersive VR-enabled devices.

VR applications used for education are often presented as either serious games or simulations (Imlig-Iten & Petko, 2018). Unlike most DGs genre, simulations are real-world representations of a system. Shute and Ke (2012) argue that, unlike simulations, games are intrinsically motivating given that they are often competitive. Nonetheless, DGs and simulations have common features like 3D models of real-world situations, educational frameworks, and contextual and interactive experiences that make them useful for learning (Imlig-Iten & Petko, 2018). VR games and simulations have been successfully used for engineering education (e.g. Beh *et al.*, 2022; Horvat *et al.*, 2022), and in most cases, emphases are not made on the type of VR application used - game or simulation. Therefore, in this chapter, VR applications and VR games are interchangeably used to refer to any interactive VR activity used for educational purposes.

Figure 2.2: Pictures of a man wearing a VR head-mounted device with hand controllers, a group of people experiencing VR via a PowerWall projection, and two people in a CAVE environment.

2.3.3 Augmented reality (AR)

Considered a type of VR, AR applications combine virtual representations with perceptions of the real world (Sherman & Craig, 2003). Whereas VR transports users to an entirely virtual environment, AR offers an interactive mix of both real and virtual worlds, as shown in figure 2.3. With AR, computer-generated information and elements are superimposed on physical environments with the help of sensors like cameras, microphones, the global positioning system (GPS) and haptic devices (Ardiny & Khanmirza, 2018).

The term AR first came to light in the 1990s following the development of AR systems for aircraft manufacturing, surgical programmes and laser printing maintenance (Cheng & Tsai, 2013). In these earlier days, AR relied on HMDs for implementation. However, with more sophisticated computing hardware and software, and advanced tracking and registration systems on these devices, mobile phones and computers are increasingly being used for AR applications (Cheng & Tsai, 2013; Pence, 2010).

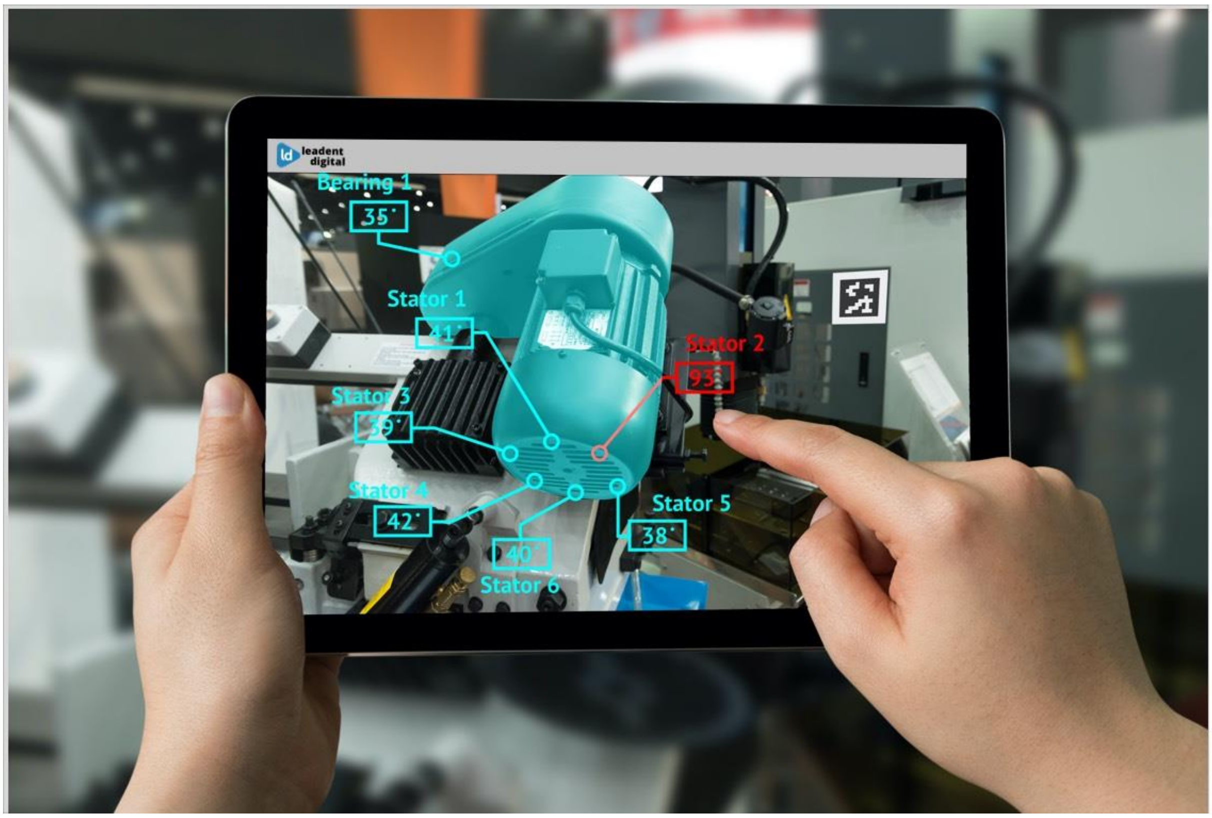


Figure 2.3: Augmented reality display on a tablet computer.

AR can be implemented as marker-based (image-based) or markerless (location-based) applications (Cheng & Tsai, 2013; Pence, 2010). In marker-based AR, specific labels are required to register the position of computer-generated objects on the real-world image, whereas in markerless AR, a wireless network or GPS is used to identify the location where the computer-generated information would be superimposed (Chen *et al.*, 2019). AR applications have been found to be very beneficial for education with great potential to positively influence the acquisition of spatial skills, practical skills and scientific enquiry (Garzón *et al.*, 2019; Cheng & Tsai, 2013).

2.4 Educational Benefits and Challenges of Immersive Technologies

Immersive technologies offer unique opportunities for students to learn through experimentation, without being exposed to the potential risks and dangers of the physical environments. Immersive technologies promote the active participation of students in learning activities through problem- and project-based learning integral to constructivism (Kiili, 2005). Interactivity built into immersive environments allows students to manipulate variables and receive immediate reactions and feedback, thus promoting experiential learning. Experiential learning focuses on the idea that learning is a process of knowledge creation through experience. According to Kolb's cycle of experiential learning, learning begins with active experimentation and concrete experience to reflective observation and abstract conceptualisation, as shown in figure 2.4 (Kolb 1984 cited in Kolb and Kolb, 2005).

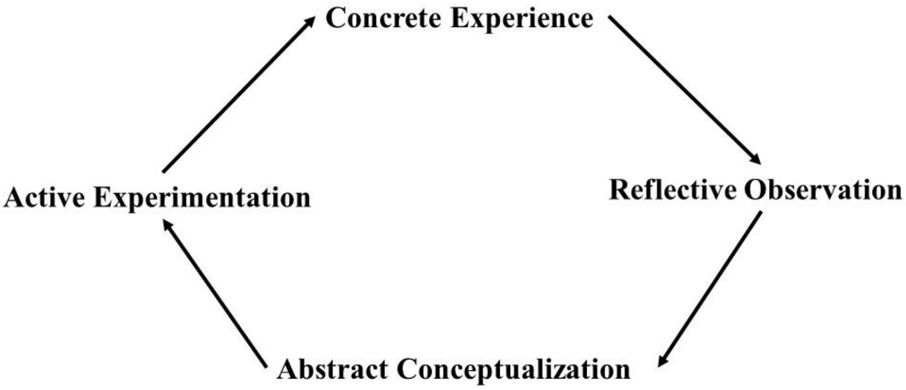


Figure 2.4: Kolb's experiential learning model (adapted from Kolb and Kolb (2005).

Immersive learning environments allow students to experience all four stages of experiential learning and create knowledge through personal experience. This is considered the basis for game-based learning (learning through games) (Kiili, 2005). Games are also known to be intrinsically motivating tools that enhance the enjoyment and time investments of students when used for L&T (Gee, 2003; Plass *et al.*, 2015). Game mechanics and incentive structures such as badges, points and leaderboards incorporated into immersive learning environments raise the interest and enjoyment of students in the learning activities (Plass *et al.*, 2015). Immersion in VR, AR and DG environments is another factor that makes these environments useful for education. In immersive learning environments, students experience a deep sense of presence in the simulated environments as they are able to interact seamlessly as they would in real life. This sense of being in the simulated environment makes learning more concrete and situated, and it is associated with longer retention (Chittaro & Buttussi, 2015). Immersive environments also facilitate flow experience which has been shown to have a positive impact on learning. Flow experience is described as “optimal experience”, the psychological state of feeling a deep sense of enjoyment, absorption and engagement in an activity (Csikszentmihalyi, 1990; Kiili, 2005). Flow experience has been found to correlate with low extraneous cognitive load (Chang *et al.*, 2017) and a higher sense of control and concentration among students in DG environments (Tsai *et al.*, 2016). For engineering education, immersive learning environments facilitate contextualised learning in a safe environment. These technologies provide safe and comparatively cheaper realistic learning environments for students to learn by experimentation. They also promote personalised and self-regulated learning commonly associated with higher academic achievements in students (Zimmerman & Schunk, 2011).

However, for formal classroom learning, there are currently some issues with the use of immersive technologies. First, high-quality HMDs are quite expensive, particularly when considering that hundreds of these would be required for a typical undergraduate engineering classroom with as many as a hundred students. This cost implication could hinder their adoption for education. VR HMDs are known to cause motion sickness also referred to as cyber or simulator sickness (Hettinger & Riccio, 1992). This can make these immersive technologies unsuitable for use by some learners. Furthermore, there are currently very few readily available relevant immersive learning applications for engineering education (Udeozor *et al.*, 2022). Designing curriculum-relevant applications requires game design skills and time commitments that only a few educators have. Without readily available applications or authoring tools that require little to no game-design skills, educators will be unable to use immersive technologies for teaching. Lastly, immersive technologies are relatively new pedagogical tools. This means that a large proportion of educators in the engineering domain may not have used these tools and lack the competency to use these for teaching. Implementation issues such as training requirements for educators to use these technologies, limited time available to integrate these technologies into classroom activities, challenges with monitoring the activities of students as well as the assessments of learning in these environments are some of the concerns raised by educators (Alyami *et al.*, 2019; Tzima *et al.*, 2019; Cooper *et al.*, 2019; Razak *et al.*, 2012). Nonetheless, with a sustained interest in immersive learning, further research in the field and advances in technology, these issues are expected to be negligible in the near future. HMDs are expected to become cheaper and better to reduce motion sickness effects on students. The development of authoring tools is increasing (Vert and Andone, 2019; Berns *et al.*, 2020) and this is expected to empower educators to develop immersive learning environments relevant to their modules. Regarding assessments, this research aims to provide educators with a design framework and best practices for measuring learning in immersive environments.

2.5 Assessment of Learning in Immersive Environments

Immersive applications can be used for formative or summative assessment, although there is a lot of support for their use for formative assessments (Kato & de Klerk, 2017; Shute et al., 2009). The assessment of learning or performance in immersive environments can be external (such as tests) or internal (game tasks) to the immersive environment. Loh, Sheng, and Ifenthaler (2015) describe these two methods of collecting data for assessment as *ex-situ* (external) and *in-situ* (internal) assessment methods. A broader description of assessment methods used with immersive learning applications was provided by Ifenthaler, Eseryel, and Ge (2012). They grouped methods of assessments with DGs as external assessment, game scoring and embedded assessment. External assessment involves the use of assessment forms such as quizzes, tests, essays and peer assessments, which are separate from the immersive environment, to determine the learning improvements or performance of students following learning activities in the environment. This assessment method is most frequently used in game-related studies due to its convenience and ease of use (Bellotti et al., 2013; C. S. Loh et al., 2015). Notwithstanding, the use of external assessment methods has been criticised as being isolated from the learning context (Groff, 2018; Shute & Ke, 2012). It is believed that using these forms of assessments misses out on the opportunity for authentic performance-based assessments, as well as the opportunity to measure more complex competencies that are otherwise difficult to measure (Groff, 2018; Shute & Ke, 2012). Game scoring, also known as monitoring of state (Hainey *et al.*, 2015) or completion assessment (Bellotti *et al.*, 2013), is an internal assessment method that focuses on measuring completed tasks, mastered obstacles, task completion times, and other achievements of the student. Game scoring provides generic summarised information on the performance of students in immersive environments with little information on the process taken to achieve such outcomes. Lastly, embedded or integrated assessment is an internal assessment method that provides detailed (grain-sized) process data on the performance of students in immersive environments. It assesses when, what, and how students interact with the environments and the choices they make. It offers valuable data for formative assessment of the learning of students. Embedded assessments are seamlessly woven into the game mechanics such that every action of the students provides indications of their competencies without interfering with gameplay (Ifenthaler *et al.*, 2012; Shute, Rahimi, and Sun, 2018). This type of information gathering is continuous, contextual, and unobtrusive, providing valid and reliable evidence of the proficiencies of students on the measured outcomes that can be used to provide relevant feedback to students (Shute *et al.*, 2018).

Although Ifenthaler *et al.* (2012) make a distinction between game scoring and embedded assessments, it can be argued that both assessment methods judge students based on their actions in the immersive environments (performance-based) rather that the answers to questions as in external assessments. The difference between these two assessment methods lies in the granularity of data gathered and not where and what kind of data is collected. For these reasons, this thesis considers external and embedded assessments as the two assessment methods investigated for immersive learning technologies. External assessment is any assessment form that is external to the immersive environment, for instance, multiple-choice questions, reports, and questionnaires. Any assessment that uses game metrics or log data of the interactions of students in the immersive environment is considered an embedded assessment. These definitions are adopted in this chapter and for the whole of the thesis. Notwithstanding the benefits of and favour for embedded assessments, external assessments offer the unique opportunity to determine conceptual understanding of demonstrated skills in immersive learning environments. Where determining conceptual understanding is crucial, performance in immersive environments might fail to provide enough information to infer deeper level understanding of desired concepts. A good balance of external and embedded assessment forms would be preferred to make accurate inferences about students' overall competencies.

Previous reviews of the literature on immersive technologies in HE showed the use of these two assessment methods with a particular preference for external assessment methods (Radianti *et al.*, 2020; Kittur and Islam, 2021; Udeozor *et al.*, 2022). Whereas these and other recent reviews have been conducted on immersive technologies for engineering education, none of these looked at all three technologies to identify trends in engineering education. The goal of this chapter is, therefore, to present a systematic review of the literature on the uses of

DGs, VR and AR in engineering education, published in the last decade. It presents current empirical research on the application of these technologies in engineering education. For this purpose, this review addresses the following research questions:

> RQ1: In what specific disciplines have immersive technologies been used for engineering education?
> RQ2: What DGs, and AR and VR devices were used in these studies?
> RQ3: What experimental design methods were used in these studies?
> RQ4: What assessment methods and outcomes were reported?

2.6 Methodology

Systematic review methodology was used to answer the research questions outlined in this chapter. Systematic reviews follow structured, transparent and comprehensive approaches to search the literature for relevant studies and provide a formal synthesis of research findings (Bearman *et al.*, 2012). Well established in the healthcare and medical domain, this methodology is associated with evidence-based practice and usually follows the international standards of the not-for-profit Cochrane organisation (Bearman *et al.*, 2012). Systematic reviews are increasingly been incorporated into educational research to uncover gaps in the literature, with discussions for best practices in place. Guided by the elaborate nine-phase process for systematic reviews for educational research by Gough (2007), the following steps were taken to ensure a transparent, structured and comprehensive systematic review process:

i. Defining review questions
ii. Defining inclusion and exclusion criteria
iii. Articulating search strategy and sources of information
iv. Screening studies to include only those meeting the inclusion criteria
v. Describing outcomes of the search strategy, typically with a systemic map
vi. Extracting relevant information from studies included
vii. Appraising the rigour or methodological quality of the studies included
viii. Synthesising all the evidence considered to answer the review questions
ix. Communicating findings and drawing conclusions

2.6.1 Digital games review

With research questions 1-4 outlined above, the first step of Gough’s review process is fulfilled. The following inclusion and exclusion criteria and search strategies fulfil steps 2 to 5:

a) the study must use a DG for engineering education;
b) the study must describe how the DG was used;
c) the study must present how students were assessed.

Studies that used non-DGs and gamification were excluded.

For all three reviews presented in this chapter, the searches were performed in five databases known to have large collections of multidisciplinary peer-reviewed journals and conference papers relevant to education and information technology. These are Web of Science, Scopus, Compendex (Engineering Village), IEEE Xplore, and Science Direct.

The search terms used include possible terms for DGs, terms synonymous with assessments and the term "engineering education". The DG terms were adopted from the review of Connolly *et al.* (2012). The search terms used for this review are:

("computer game" OR "video game" OR "serious game" OR "simulation game" OR "game-based learning" OR "online game") AND ("assessment" OR "feedback" OR "performance" OR "effectiveness") AND ("engineering education").

Several elimination steps were carried out as the search engines returned a large number of articles that did not fully meet all the inclusion criteria. As presented in figure 2.5, a total number of 3032 papers from all five databases were initially returned by the search engines but only 20 papers were selected as meeting the inclusion criteria.

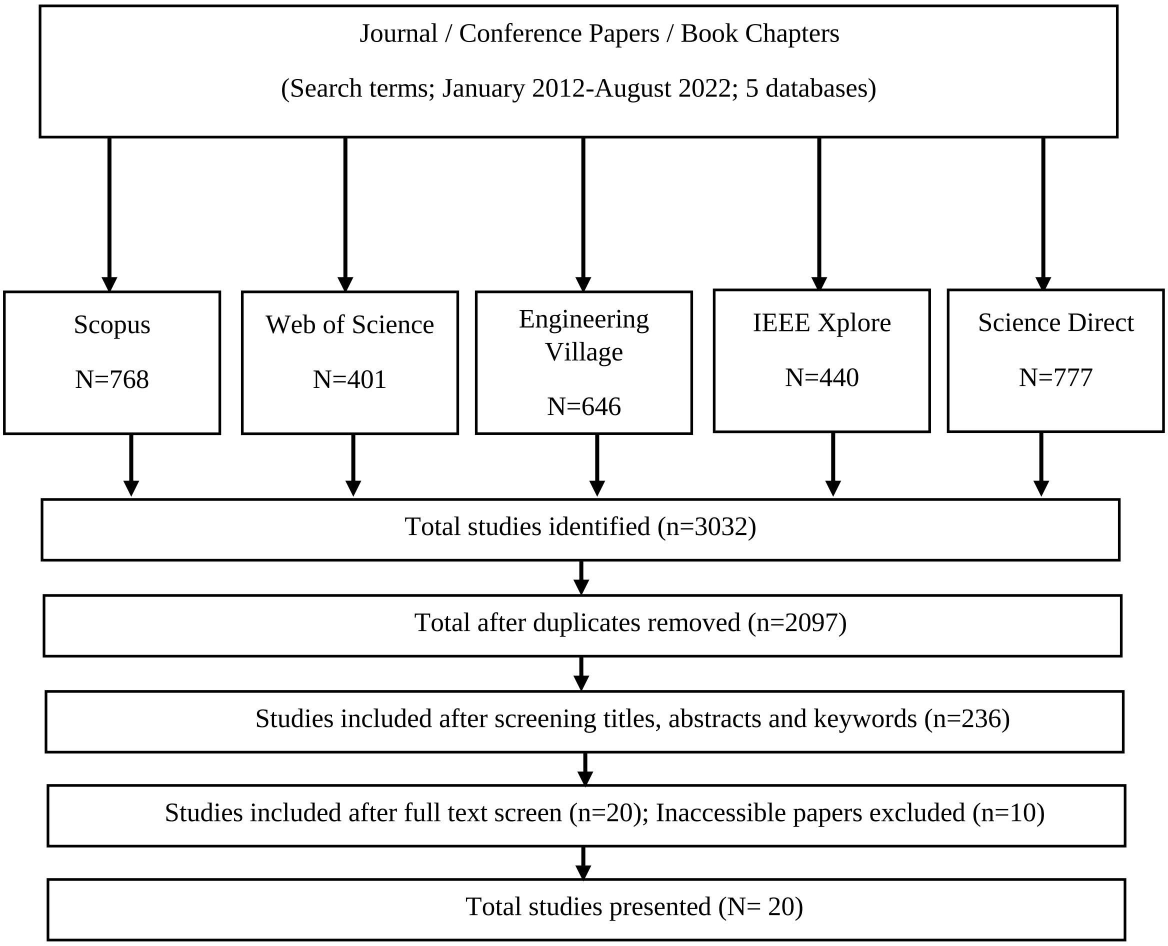


Figure 2.5: Systematic review process of publications on DGBL in engineering education from 2012 to 2022.

The distribution of the papers on the use of DGs for engineering education identified in this review by the year and the medium of publication is shown in figures 2.6 and 2.7, respectively. The majority of the papers included in this review (n=13, 65%) are journal articles. The highest number of publications on DGs was seen in 2018 and 2020. Ten papers that could not be accessed via the university library database or any open database were excluded from the review. A list of the journals, conference proceedings and books where these papers were published is presented in Appendix 2. A.

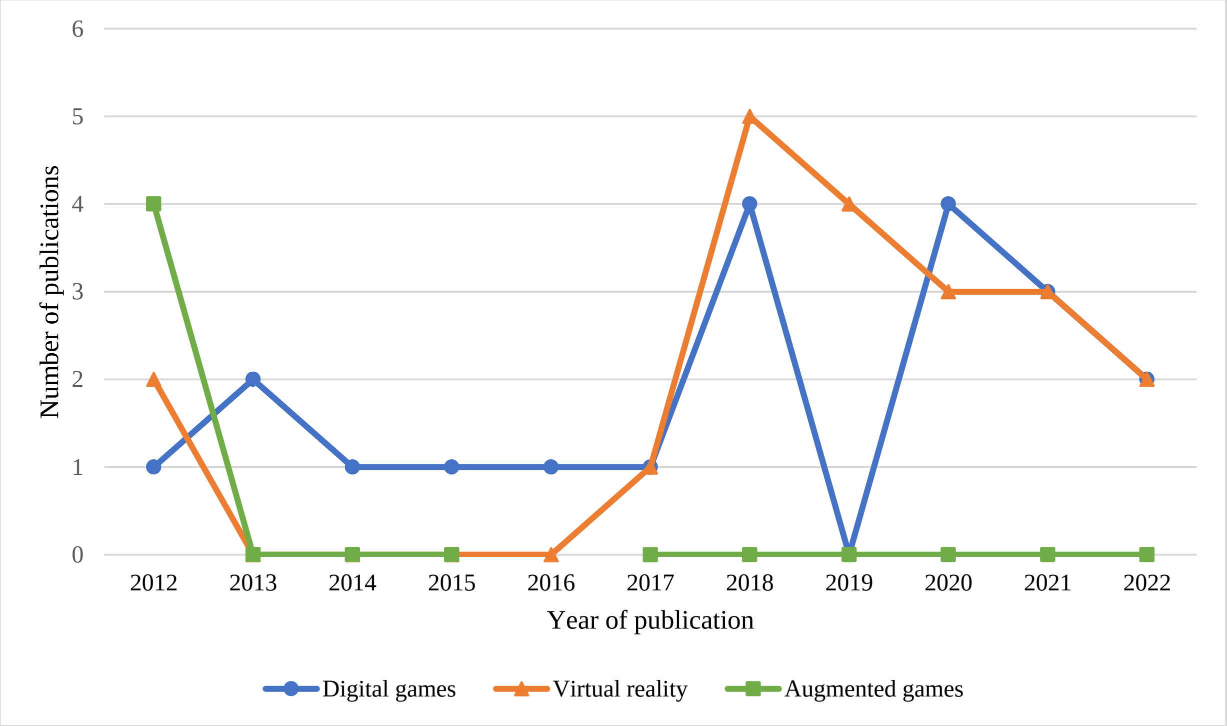


Figure 2.6: Distribution of DG, VR and AR papers by year of publication.

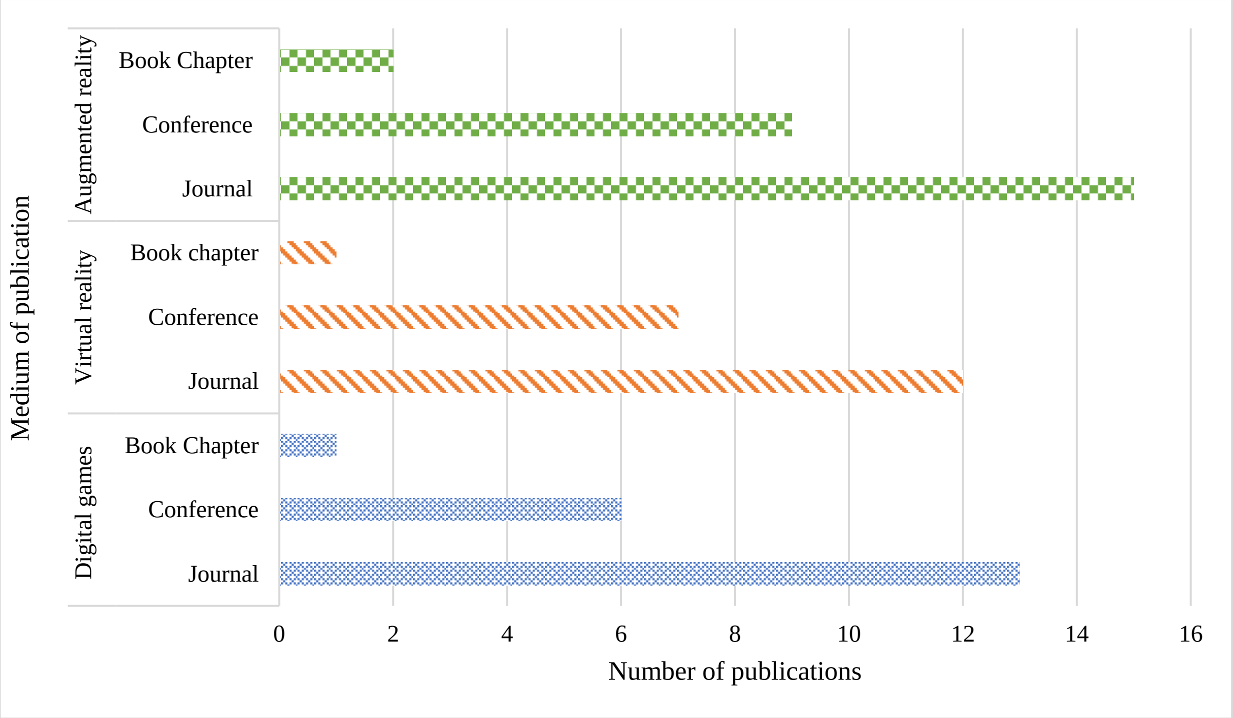


Figure 2.7: Distribution of DG, VR and AR papers by the medium of publication.

2.6.2 VR review

The search criteria used for the VR literature review are:

a) the study must use an immersive VR application accessed using HMDs, CAVE or PowerWalls for engineering education;

b) the study must describe how the immersive VR was used;

c) the study must indicate the VR devices used;

d) the study must present how students were assessed.

Studies that described virtual laboratories on desktops were excluded. The search terms used for this review are:

(“virtual reality” AND “engineering education”) AND (“assessment” OR “feedback” OR “performance” OR “effectiveness”).

The search on all five databases returned a large number of articles that did not fully meet all the inclusion criteria. From a total number of 1969 papers, only 20 papers were selected as meeting the inclusion criteria, as shown in figure 2.8.

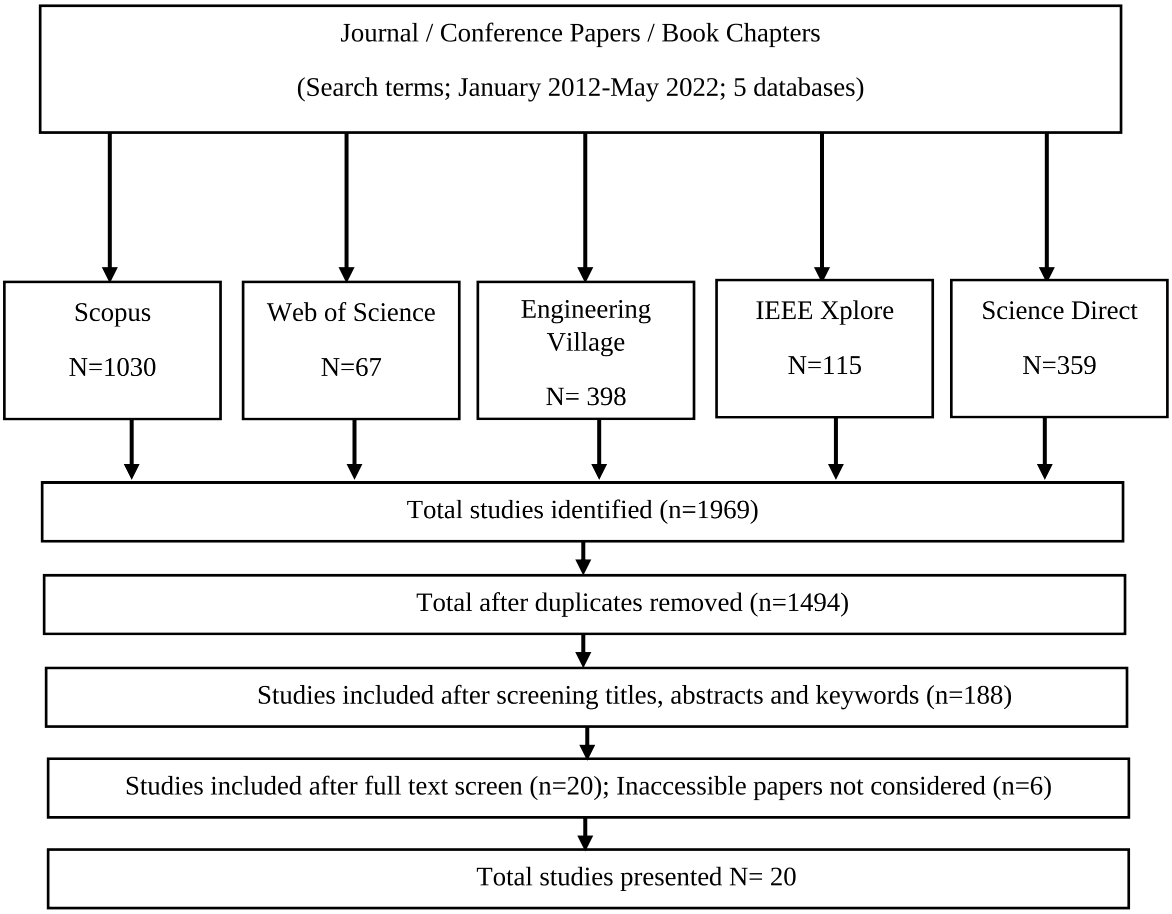


Figure 2.8: Systematic review process of publications on VR in engineering education from 2012 to 2022.

The distribution of the papers on the use of VR for engineering education identified in this review by the year and the medium of publication is shown in figures 2.6 and 2.7. The majority of the papers included in this review (n=12, 60%) are journal articles. The highest number of publications on the subject was seen in 2018 (n=5, 25%). The journals, conference proceedings and books where these studies were published are presented in Appendix 2. B.

2.6.3 AR review

The search criteria used for the AR literature review are:

a) the study must use an AR application for engineering education;

b) the study must describe how the AR application was used;

c) the study must present how students were assessed.

The search terms used in this review are:

("augmented reality" AND "engineering education") AND ("assessment" OR "feedback" OR "performance" OR "effectiveness").

Several elimination steps were also carried out to limit the large number of articles returned by the search engines to only those papers that meet the inclusion criteria. From a total number of 994 papers from all five databases, only 26 papers were selected as presented in figure 2.9.

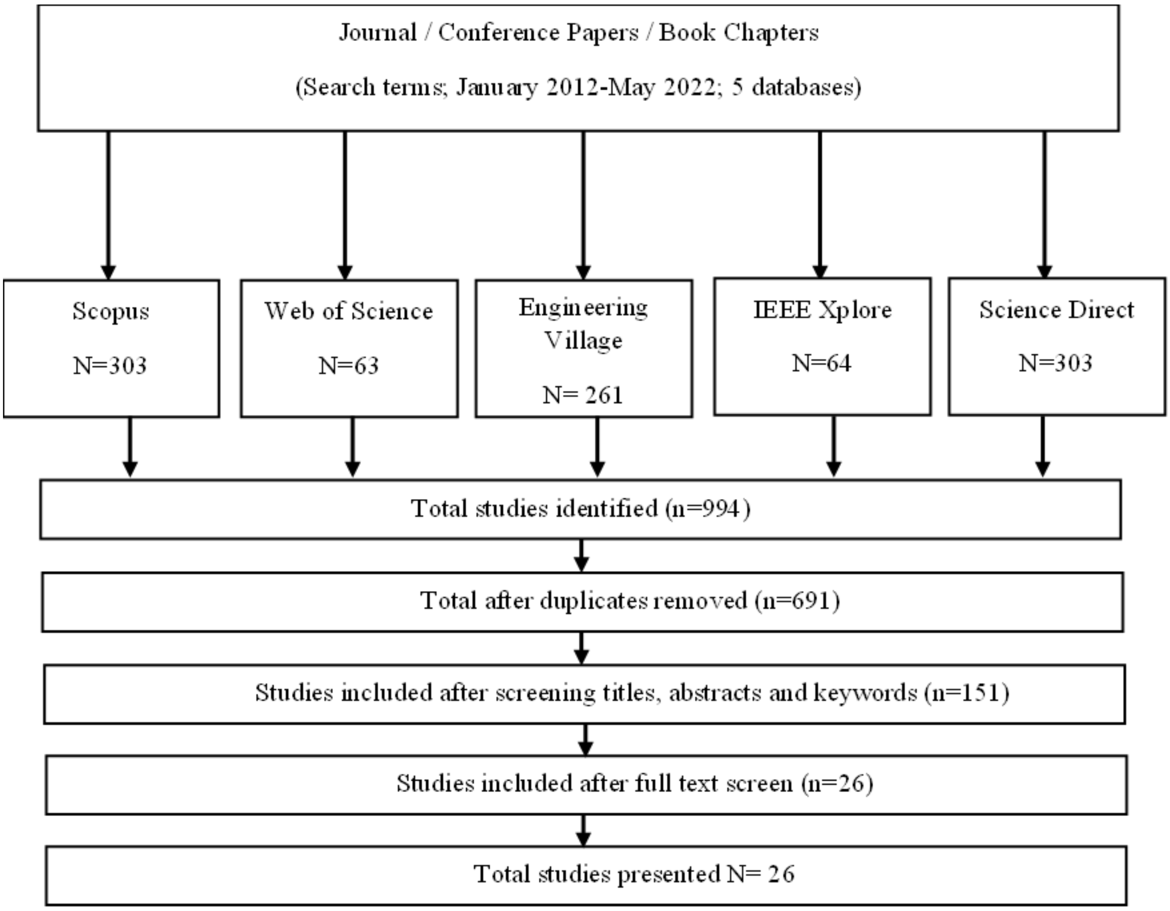


Figure 2.9: Systematic review process of publications on AR in engineering education from 2012-2022.

The distribution of the papers on the use of AR for engineering education identified in this review by the year and the medium of publication is shown in figures 2.6 and 2.7. More than half of the papers included in this review (n=15, 58%) are journal articles. The highest number of publications on this subject was seen in 2012 and 2019 (n=4 in each; 15%). The journals, conference proceedings and books of publication of the reviewed papers are presented in Appendix 2. C.

2.7 Data Analysis

All papers included in this systematic review presented data to answer the research questions outlined in section 2.5 of this chapter. Table 2.3 presents a breakdown of the relevant information obtained from the studies presented in the reviewed papers in fulfilment of step 6 of Gough's systematic review process.

Table 2.3**:** Relevant information from the systematic review of DGs, VR and AR for engineering education.

Authors and Year	**Research Design**	**Assessment Methods**	**Sample Size**
Digital Games			
(Anupam *et al.*, 2018)	Time series	Pre/post tests	14
(Cardin *et al.*, 2015)	Randomised; correlational	In-game data Questionnaire	46
(Chang *et al.*, 2012)	Time series	Pre/post tests	25
(Chang *et al.*, 2016)	Quasi-experiment	Pre/post tests Final exams	110
(Dib & Adamo-Villani, 2014)	Randomised experiment	Pre/post tests	42
(Duin *et al.*, 2013)	Time series	Pre/post self-assessments questionnaires	24
(Flores *et al.*, 2020)	Randomised experiment	Pre/post tests Questionnaire	20
(Miljanovic & Bradbury, 2020)	Quasi-experiment	Questionnaire Log data	100
(Perini, Luglie*tti*, *et al.*, 2018a)	Quasi-experiment	Pre/post tests	62
(Perini, Oliveira, *et al.*, 2018)	Time series	Pre/post tests	33
(Shernoff *et al.*, 2020)	Quasi-experiment	Pre/post tests	243
(Siddique *et al.*, 2013)	Randomised experiment	Post tests	66
(Smith & Chan, 2017)	Quasi-experiment	Pre/post tests	485

		Mid and Final exams	
(Suescún *et al.*, 2018)	Quasi-experiment	Post tests Exams	37
(Xenos & Velli, 2020)	Time series	Pre/post tests	144
(Zhao *et al.*, 2021)	Time series	Pre/post tests Questionnaire	87
(Baruah & Mao, 2021)	Time series	Questionnaire	80
(Pando Cerra *et al.*, 2022)	Randomised experiment	Post tests Questionnaire	96
(Evangelou *et al.*, 2021)	Time series	Pre/post tests Questionnaire	71
(Gordillo *et al.*, 2022)	Quasi-experiment	Pre/post tests Questionnaire	180
Virtual Reality			
(Peng *et al.*, 2012)	Time series	Questionnaire	21
(Chaturvedi *et al.*, 2012)	Quasi-experiment	Tests Questionnaire	39
(Im *et al.*, 2017)	Time series	Questionnaire	26
(Jaksic, 2018)	Time series	Tests, projects, questionnaire	Unknown
(McCusker *et al.*, 2018)	Quasi-experiment	Post-tests	45
(Ostrander *et al.*, 2018)	Randomised experiment	Pre/post tests	44
(Akbulut *et al.*, 2018)	Quasi-experiment	Post-tests	36
(Ritter *et al.*, 2018)	Quasi-experiment	Pre/post tests	196
(Bashabsheh *et al.*, 2019)	Time series	Questionnaire	22
(Mirabolghasemi *et al.*, 2019)	Not described	Tests and homework	10
(Salah *et al.*, 2019)	Randomised experiment	Game metrics Questionnaire	50
(Lin & Wang, 2019)	Randomised experiment	Pre/post tests	30
(Kandi *et al.*, 2020)	Repeated measures	Tests	94
(Ostrander *et al.*, 2020)	Randomised experiment	Pre/post tests Questionnaire	107
(Halabi, 2020)	Quasi-experiment	Report, work product, presentation, questionnaire	67
(Bolkas *et al.*, 2022)	Quasi-experiment	Graded work, exams, questionnaire	21

(Rossado Espinoza *et al.*, 2021)	Not described	Pre/post tests	42
(Kamińska *et al.*, 2021)	Randomised experiment	Pre/post tests Questionnaire	117
(Beh *et al.*, 2022)	Quasi-experiment	Game metrics Tests	30
(Horvat *et al.*, 2022)	Quasi-experiment	Work product	40
Augmented Review			
(Martin-Gutierrez *et al.*, 2012)	Time series	Delayed hands-on tasks	20
(Roca González *et al.*, 2012)	Quasi-experiment	Pre/post tests	91
(Sánchez *et al.*, 2012)	Quasi-experiment	Pre/post tests	25
(Contero *et al.*, 2012)	Quasi-experiment	Pre/post tests	42
(Tixier & Albert, 2013)	Single subject (multiple baselines)	Pre/post tests	39
(Martín-Gutierrez *et al.*, 2013)	Quasi-experiment	Pre/post tests	250
(Fon*seca et al.*, 2014)	Quasi-experiment	Exam scores	57+
(Gutiérrez & Fernández, 2014)	Quasi-experiment	Exam scores	47
(Shirazi & Behzadan, 2015)	Quasi-experiment	Work product	60
(Martín-Gutiérrez *et al.*, 2015)	Quasi-experiment	Pre/post tests	49
(Frank & Kapila, 2017)	Time series	Pre/post tests	75
(Alrashidi *et al.*, 2017)	Quasi-experiment	Pre/post tests Questionnaire Game metrics	20
(Turk*an et al.*, 2017)	Quasi-experiment	Pre/post tests	41
(Luo & Mojica Cabico, 2018)	Randomised experiment	Pre/post tests	40
(Guo, 2018)	Unknown	Pre/post tests	32+
(Bairaktarova *et al.*, 2019)	Unknown	Pre/post tests	119
(S*ingh et al.*, 2019)	Randomised experiment	Pre/post tests	60
(Veide & Strozheva, 2019)	Unknown	Pre/post tests	72
(Reut*er et al.*, 2019)	Randomised experiment	Questionnaire Work product	14
(Gómez-Tone *et al.*, 2020)	Randomised experiment	Pre/post tests	171
(Urbano *et al.*, 2020)	Quasi-experiment	Pre/post tests	433
(Guo & Kim, 2020)	Unknown	Post-tests	45
(Criol*lo-C et al.*, 2021)	Randomised experiment	Post-tests	80
(Urbina Coronado *et al.*, 2022)	Unknown	Post-tests	30

(Kearney *et al.*, 2022)	Randomised experiment	Pre/post tests	117
(McCord *et al.*, 2022)	Unknown	Pre/post tests Questionnaires Observations	53

+ papers reported the sample sizes of some participating students but not all.

2.8 Findings and Discussions

Steps 7-9 of Gough's review process are presented in the results section below.

2.8.1 Engineering disciplines identified

Ten distinct engineering disciples shown in figure 2.10 were identified in the reviews. Studies that used immersive learning technologies in computer (n=6), electrical and electronics (n=6), industrial (n=5), and mechanical (n=13) engineering disciplines were identified in this review with the mechanical engineering discipline having the highest number of publications. A total of 19 papers either failed to mention the specific engineering disciplines where the immersive technologies were used (*e.g.* Criollo-C *et al.*, 2021; Horvat *et al.*, 2022; Pando Cerra *et al.*, 2022) or used the technology in more than one engineering discipline (*e.g.* Beh *et al.*, 2022; Reuter *et al.*, 2019). These 19 papers were assigned to a "general engineering" category.

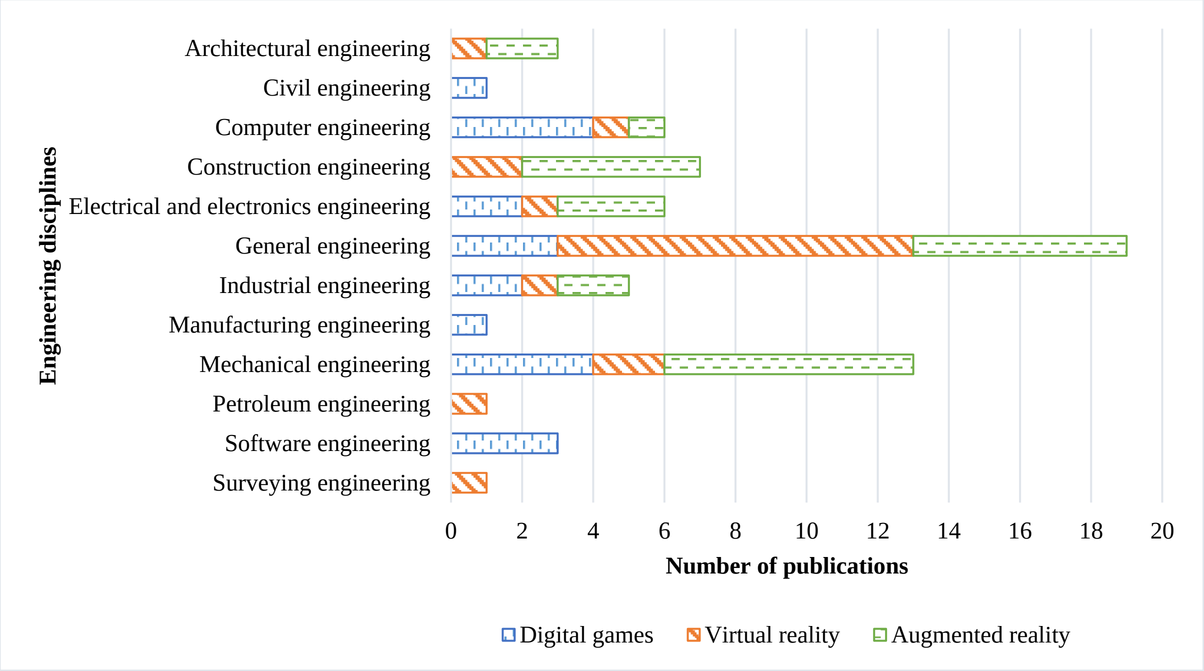


Figure 2.10: Engineering disciplines identified in the review of (n=66) papers published from 2012 to 2022 in five databases: Scopus, Web of Science, Engineering Village, IEEE Xplore, Science Direct.

A range of competencies was targeted in the studies, from core engineering knowledge and skills such as nanotechnology (Akbulut *et al.*, 2018), well trajectories (Mirabolghasemi *et al.*, 2019), internet protocol (IP) addressing and ISO/OSI model (Criollo-C *et al.*, 2021), to skills like decision making (Cardin *et al.*, 2015) and ethics (Xenos & Velli, 2020).

Given that the current review is limited to studies that presented the assessments of learning with immersive technologies, the findings show that of all ten engineering disciplines identified, mechanical engineering discipline has shown more advancement with the use of these technologies. Generally, with new technologies, the evaluation of prototypes and user experiences is first conducted before being launched and used for the intended purpose. It is often the case that the majority of engineering studies present prototypes and user experience evaluations with only a few going beyond these to measure effectiveness (Udeozor *et al.*, 2022).

AR happened to be the most popular immersive technology identified in this review with 26 reviewed papers, while DGs and VR had 20 papers each. The affordances and low cost of some AR-enabled devices are possibly fuelling the wider adoption of AR technology for engineering education compared to VR. The next sections will highlight other likely reasons for the higher percentage of AR papers in this review.

2.8.2 Digital games and VR/AR devices used

Ninety percent (90%) of the DG papers included in this review described the type of games used in their studies. Some of the papers utilised more than one game (or game versions) in their studies (*e.g.* Miljanovic and Bradbury, 2020; Zhao *et al.*, 2021). All the games reportedly used are educational games presented in table 2.4. Unlike digital entertainment games (DEGs), educational and serious games are designed with specific learning outcomes in mind. For a technical domain, it is understandable that the DGs used were specially designed to teach engineering-relevant competencies. Using an existing DEG game for core engineering education would be challenging. However, in the absence of curriculum-aligned DGs, engineering students risk missing out on the learning experiences DGs provide. This finding emphasises the need for more curriculum-aligned DGs, VR and AR applications to be made readily available to educators in engineering faculties. For more general transferrable skills in non-engineering domains, several other studies have successfully used DEG games (Adams *et al.*, 2016; Shute *et al.*, 2015).

Table 2.4: DGs identified in the review of publications on engineering education from 2012 to 2022.

Disciplines	**Digital Games**
Civil Engineering	Sustainability Challenge game
Computer engineering	Loop game; Function game; Structure game EEEE—Expert and Efficient Estimators Enterprise My life as a software engineer
Electrical engineering	SimVenture Evolution ‘Particles in a Box’ game
General Engineering	Space Race DIBROOM
Industrial Engineering	Life Cycle Assessment game Sustainable Global Manufacturing game
Manufacturing engineering	Life Cycle Assessment game
Mechanical engineering	Garry’s Mod game – virtual adaptation Mecagenius game Race Car game Gear train virtual lab game Spumone
Software engineering	SimulES-W; Ethical Dilemmas GidgetML; Gidget

The devices on which these immersive applications were accessed by engineering students are presented in figure 2.11. Computers were mostly used for DGs while HMDs were most popular for VR. Computers (including tablets) and mobile devices were the two devices mainly used by students for AR. It is worth reiterating that this review did not include VR papers that used only computers but limited its search to papers reporting the use of immersive VR tools such as HMDs, PowerWalls and CAVE technologies. Hence, figure 2.11 should not be viewed as a comparison of devices used for DGs, VR and AR, but as a visual representation of the devices reportedly used in the reviewed papers.

Immersive VR-enabled devices are currently expensive, but their costs are expected to reduce over time. However, of the three identified VR devices, HMDs are the cheapest at the moment (Havig *et al.*, 2011). This is possibly the reason for their wider use among the papers reviewed. Interestingly, of the HMDs available on the market, the Google Cardboard which is considered the cheapest option there is was used in only one of the 20 reviewed VR papers (Akbulut *et al.*, 2018). The other HMDs used include Oculus Rift (n=6), HTC Vive (n=7), Samsung Gear (n=1) and one unknown HMD. Oculus Rift and HTC Vive are two of the most expensive and top-rated HMDs currently available (Mehrfa*rd et al.*, 2019). Their use for engineering education is reassuring as it suggests that institutions see the value of immersive technologies thus investing in higher-quality devices that would potentially result in better learning experiences.

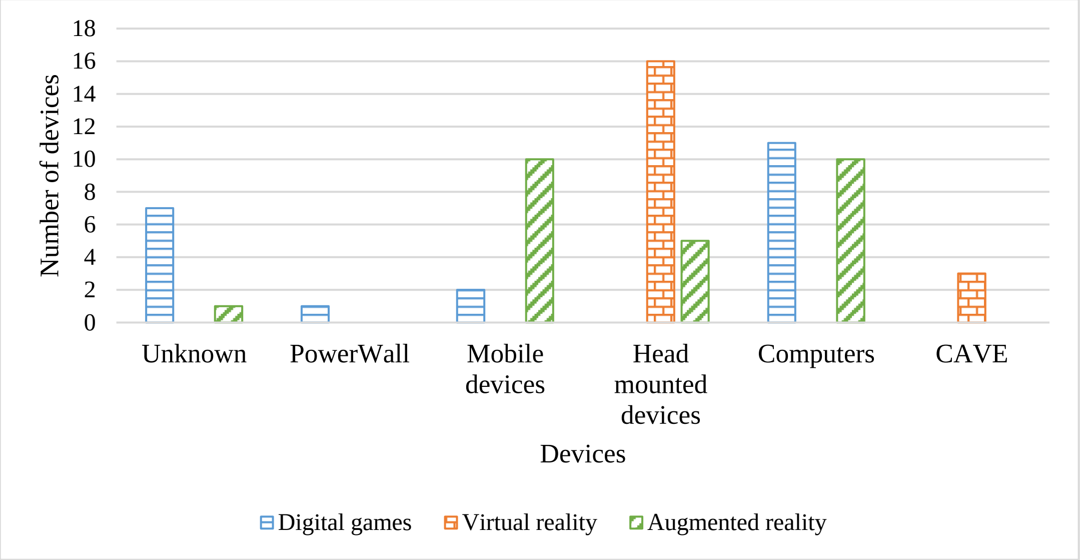


Figure 2.11: Devices reportedly used in the (n=66) reviewed papers on the use of DGs, VR and AR for engineering education published between 2012 and 2022.

Mobile devices and computers were the most frequently used AR-enable devices identified in this review; a trend that was also found in previous systematic reviews (Cheng & Tsai, 2013; Ibáñez & Delgado-Kloos, 2018). Whereas earlier versions of AR required HMDs, recent applications of AR are exploiting the advances in mobile technologies, using them as alternatives to expensive, obtrusive and cumbersome AR HMDs (Kerawalla *et al.*, 2006; Wu *et al.*, 2013; Garzón, Pavón and Baldiris, 2019). Compared to the studies in the VR review, studies in the AR reviews reported larger sample sizes with as many as 433 participants in one (Urbano *et al.*, 2020). This shows the potential for the wider adoption of AR in HE given that most of the students have mobile devices, and in many institutions, hand-held computers are distributed to students in addition to the large computer clusters available to them. The increasing use of mobile AR and their abilities to provide similar or better AR learning experiences is promising and a step forward in the move toward experiential learning in engineering education.

2.8.3 *Experimental designs used*

Several experimental research designs were used in the studies reviewed to assess the effectiveness of immersive technologies for engineering education. Sixteen studies (24% of all the reviewed papers) used randomised assignments to place students into treatment and control groups. This random assignment of participants makes this type of experimental design the most rigorous of all experimental designs and is commonly referred to as “true experiment” (Creswell, 2011). In educational research where it is not always physically possible to randomly assign students into treatment and control groups, alternative design methods, such as quasi-experiments, are used. Quasi-experimental designs were used by the largest proportion of studies in this review as shown in figure 2.12.

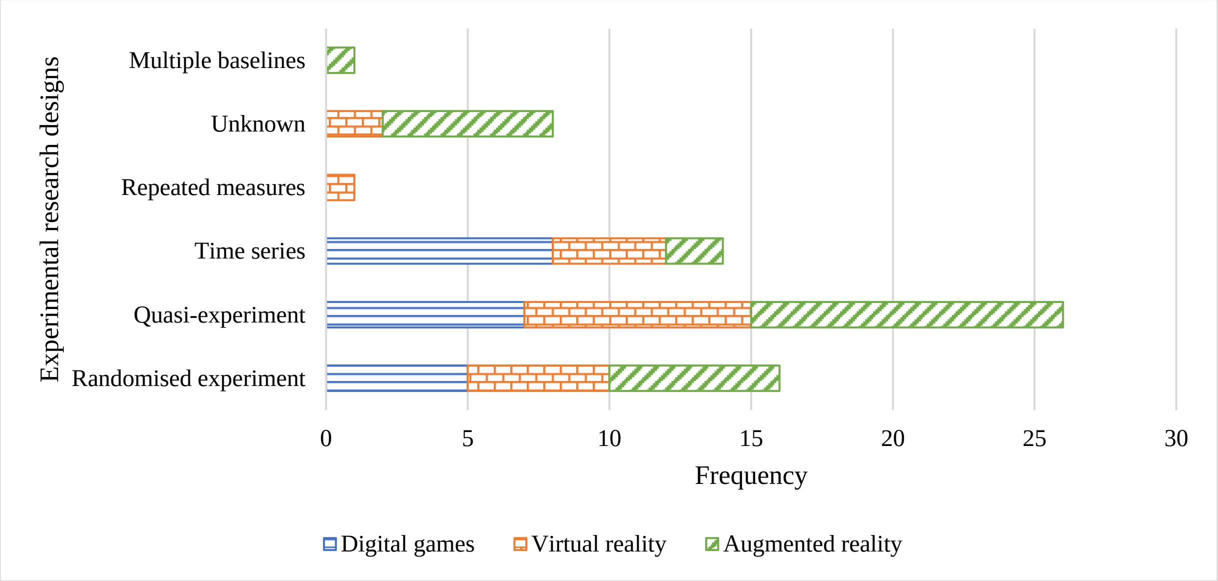


Figure 2.12: Experimental designs used in the (n=66) reviewed papers on DGs, VR and AR for engineering education published between 2012 and 2022.

Quasi-experimental design involves a non-randomised assignment of participants into groups. Both randomised and quasi-experimental designs fall into the “between-group designs” that enable group comparisons (Creswell, 2011). In some of the reviewed papers, students in two classrooms were used, with the treatment group assigned to one classroom and the control group to the other (Akbulut, Catal and Yıldız, 2018; Perini, Luglietti, *et al.*, 2018). Others simply asked volunteers to choose to be in the game group or control group (Smith & Chan, 2017). These actions often lead to selection bias which is the result of inherent differences between participants in experimental and control groups (Gopalan *et al.*, 2020). The outcomes of the experiments could also be influenced by the characteristics of students in the groups. Randomised and quasi-experimental designs are most useful for comparing groups and ascertaining the effectiveness of an intervention, while the other designs are best suited for smaller sample sizes as highlighted in table 2.5.

The remainder of the papers used either single-group designs or did not specify the experimental design used. Time series, multiple baselines and repeated measures designs (see below) are single-group experimental designs used when there is a limited number of participants and in other cases where it is impossible or unethical to administer an intervention to a select few, such as in randomised and quasi-experimental designs (Creswell, 2011). Of the three single-group experimental designs identified, time series had the highest number of papers included in this review. In time series designs, a single group is studied over a period of time with several test measures or observations conducted during the period (*e.g*. Chang *et al.*, 2012; Xenos and Velli, 2020; Zhao *et al.*, 2021). As outlined in table 2.5, time series design is used when the participants of a study are few. While it can be used to measure changes after an intervention, this design is labour-intensive and cannot be used to determine how an intervention compares to another (Creswell, 2011). One VR study utilised repeated measure design, an experimental design that requires all research participants of a study to take part in all experimental treatments, and performances under these treatments are compared (*e.g.* Kandi *et al.*, 2020). This experimental design is very useful for comparing the effects of different interventions on participants. However, outcomes observed with repeated measures designs can be influenced by the exposure of participants to previous interventions (Creswell, 2011). Multiple baseline design was used in one of the AR studies to determine the effect of AR on hazard recognition at various levels of intervention (Tixier & Albert, 2013). This experimental design, involving administering interventions or treatments to each participant at different times is often used to prevent treatment diffusion among participants (Creswell, 2011). Time series, repeated measures and multiple baseline experimental designs fall under the “within-group designs” (Creswell, 2011). These research designs have their strengths and weaknesses, and the choice of which to adopt would depend on factors such as the goal of the study, access to participants, sample size of participants, and available resources, to name a few. In an ideal world, when the goal is to compare or determine the effectiveness of an intervention, randomised experiments offers the best controls over threats to validaty, and thus should be adopted.

2.8.4 Assessment methods and outcomes

To explore the effectiveness of DGs, VR and AR for engineering education, the reviewed papers employed a variety of methods shown in figure 2.13. 32% of the papers in this review used more than one assessment method which typically involved the use of tests and questionnaires. External assessment methods such as pre-and/or post-intervention (DGs, VR, AR) tests made up more than half (55%) of all assessment types used in the studies in this review.

Table 2.5: Characteristics, advantages and disadvantages of the experimental designs identified in the review of (n=66) papers on DGs, VR and AR for engineering education.

Experimental designs	**Characteristics**	**Advantages**	**Disadvantages**
Randomised experiment	Participants are randomly selected and assigned to groups. Research environments are often controlled. Allows for the manipulation of experimental conditions. One or more interventions are used	Limited threats to validity.	Controlled environments might influence the behaviours of participants. Sometimes unethical to apply to educational research.
Quasi-experiment	Non-randomised selection of participants; participants are selected based on relevant characteristics. Natural environments are often used.	An alternative to randomised experiments and useful for group comparisons in educational research.	Higher threats to validity due to biases inherent in the compared groups. Limited controls over external influencing factors.
Time series	One group studied over a given period. Used when it is impractical to administer interventions to a select few.	Useful when the number of participants is low Easy to control threats to internal validity.	Design is labour-intensive.
Repeated measures	Single group studied Performance under an intervention is compared with performance under another intervention.	Requires smaller sample sizes	Performance in one intervention can be influenced by exposure to previous interventions known as order effect
Multiple baselines	Single-subject or group designs are used to administer interventions at different times to each participant.	Better control for threats to internal validity.	Time-consuming and resource intensive.

Self-assessment questionnaires, delayed tests such as exams, and others such as reports and presentations, made up 24%, 8%, and 8% of the assessment types in this review, respectively.

95% of the assessment methods used by the studies reviewed are external assessments. Only 5% of the assessments used were embedded assessments.Although unsurprising given previous findings (Radianti *et al.*, 2020; Kittur and Islam, 2021), this shows that the full potential of immersive technologies for authentic assessment is still far from being explored. Bellotti *et al.* (2013) and Loh *et al.* (2015) explained that the preference for external assessment methods is due to the challenges with designing, collecting and analysing embedded assessments. These challenges highlighted in table 2.6 are likely to limit the adoption of embedded assessments given that almost a decade since their papers, it is still the case that external assessments are predominantly used in these studies. This certainly brings to light the need for a convenient and user-friendly method of designing and implementing embedded assessments when using immersive learning technologies.

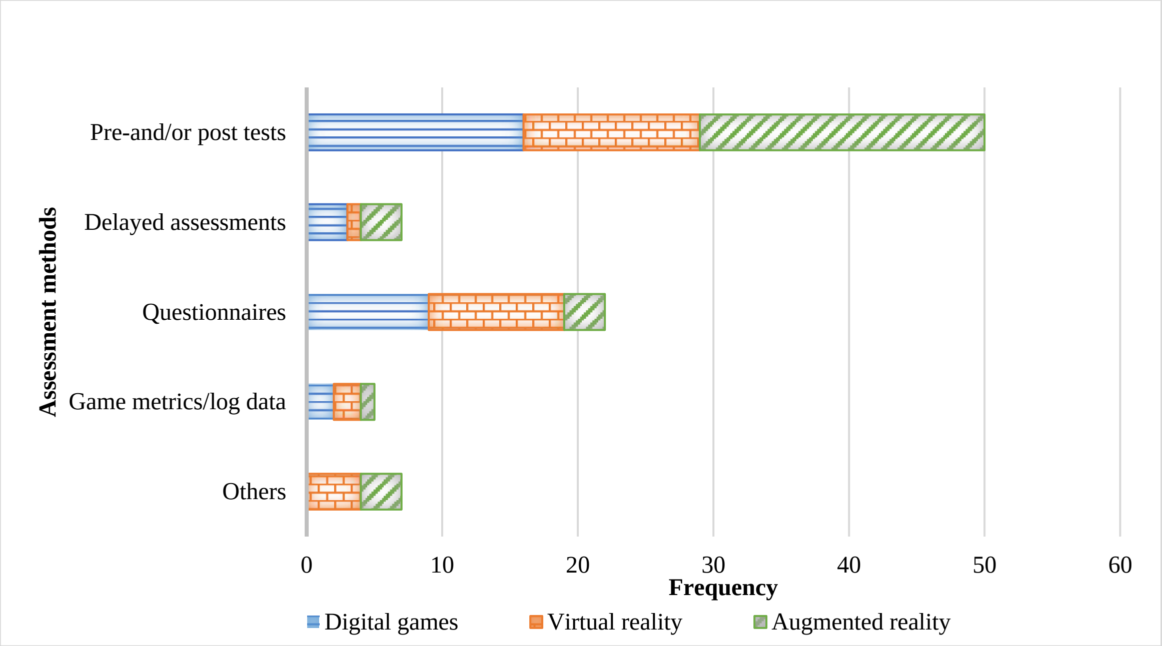


Figure 2.13: Assessment methods identified in the reviewed (n=66) papers.

The findings of the studies in this review suggest that immersive technologies enhance learning among engineering students. In the study of Suescún *et al.* (2018), a DG was used to teach software engineering concepts to 37 engineering students. Using quasi-experimental design, it was found that overall, both game and non-game groups had comparative mean post-test scores (Mean_{DG} = 7.89, s.d. = 1.87; $\text{Mean}_{\text{ctrl}}$ = 7.44, s.d = 3.7). Moreover, on a delayed test, it was found that students in the game group performed better than those in the control group (Mean_{DG} = 7.26, s.d. = 1.3; $\text{Mean}_{\text{ctrl}}$ = 6.33, s.d. = 2.33) with 64% of students in the DG group scoring above 7 on related exam questions compared to 47% from the non-game group. Similarly, Xenos and Velli (2020) used a DG to teach ethics to (n=144) software engineering students.

Table 2.6: Strengths and limitations of assessment methods relevant to immersive learning.

Assessment methods	Assessment types	Strengths	Limitations
External	Pre-and/or post-tests	Familiar to educators. Easy to design. Provides immediate indications of the effectiveness of an intervention.	Less useful for authentic assessments. Less effective for skills and performance-based assessment that measures knowledge applications and procedural skills in real-world contexts. Time-consuming to implement tests before and after an intervention.
	Delayed tests	Easy to design. May not require the design of another assessment besides classroom exams. Provides evidence of long-term retention of the competence learned.	Less useful for performance and skills-based assessments that measure knowledge applications and procedural skills in real-world contexts.
	Questionnaires	Easy to design and implement. Provides information about the opinions of learners.	Subjective data are prone to biases. Not useful for measuring learning effectiveness.
Embedded	Log data	Provides performance-based evidence of learning. Facilitates authentic assessment. Provides process data of actions of learners that can be used for formative assessments.	Challenging to design embedded assessments Log data can be too large and contain unnecessary information. Cleaning and processing log data is labour-intensive Analysing log data often requires skills in the use of sophisticated statistical tools.

The results showed that the post-game test scores of students were significantly higher than their pre-test scores ($\text{Mean}_{\text{pre-test}}$ = 67.99, s.d.= 16.92; $\text{Mean}_{\text{post-test}}$ = 80.59, s.d.=15.57; Z=8.321, *p*<.0001). Using the Analysis of Covariance (ANCOVA) test, Lin and Wang (2019) found statistically significant differences between the post-test performance of 30 students who were randomly assigned to learn with VR and those who were assigned to learn with traditional learning materials (Mean_{VR} = 32.4; s.d.=3.4; $\text{Mean}_{\text{ctrl}}$=37.2; s.d.=2.3), with a large effect size (Cohen's d = 0.637). Im *et al.* (2017) used a self-assessment questionnaire to measure the perceived learning effectiveness of VR by (n=26) engineering students. The result showed that 70% of the students perceived VR to be effective for learning about systems architecture (Mean = 3.85, s.d. =1.0).

One study used exam scores of a cohort of students who used an AR application for learning representation systems as part of their architectural studies, to compare to the score of previous cohorts who only learned using traditional approaches (Fonseca *et al.*, 2014). A significant increase was found between the average exam scores of the students who used AR compared to those of previous cohorts (Mean_{AR} = 6.54, s.d.= 0.1; $\text{Mean}_{\text{ctrl}}$ = 5.62, s.d.=0.03; *p*<0.05). These outcomes were attributed to the use of immersive technologies, thus implying that immersive technologies can be effective for engineering education. However, generalising the outcomes of these studies would require a critical evaluation of the research methods, sample sizes and statistical rigour of each of the studies. It is not the aim of this chapter to judge the quality of the studies in this review, nor to generalise the outcomes of the assessment of the effectiveness of immersive technologies. The goal is to provide an overview of the current pedagogical practices with DGs, VR and AR in engineering education in order to inform their use for teaching and assessment.

It is important to note that while positive outcomes were reported in most of the studies, some studies found that immersive technologies were no better than traditional teaching methods (Reuter *et al.*, 2019; Beh *et al.*, 2022). The performance of a few engineering students (n=14) who were randomly assigned to AR (using HoloLenses) and desktop 2D groups was compared using scores on a 100-credit task (Reuter *et al.*, 2019). The results showed the 2D desktop group scored higher on average compared to the AR group (Mean_{AR} = 68.21, s.d.= 13.49; $\text{Mean}_{\text{ctrl}}$ = 79.43, s.d.=9.74; *p*>0.05) but the differences were not statistically significant, and the cohort sample size was very small. In the study of (n=30) engineering students by Beh *et al.* (2022), it was also found that students in the control group were faster at completing given tasks on paper compared to the VR group that was required to complete similar tasks in VR ($\text{Time}_{\text{ctrl}}$ = 13.60 minutes; Time_{VR} = 24.37 minutes). These outcomes can be attributed to the use of VR for tasks that are usually performed using traditional methods, especially given that the VR students performed better in a multiple-choice test compared to the control group ($Mean_{VR} = 4.07$, s.d.= 0.7; $Mean_{ctrl} = 3.93$, s.d. = 0.8). Nevertheless, these outcomes have implications for the use of immersive learning technologies for formal classroom learning and assessments.

In another study, it was found that students who learned manual material handling with AR performed comparatively worse than students who learned the traditional way (Guo & Kim, 2020). Guo and Kim compared the workload and performance of students who learned with AR and those who learned with traditional materials and found a significantly higher mental workload in AR students compared to the control group ($Mean_{AR}= 62.66$, s.d.= 11.81; $Mean_{ctrl} = 49.62$, s.d.= 15.34; $p<0.05$). The performance of the AR group on the given tests was also lower than that of the control group ($Mean_{AR} = 58.13$, s.d. = 22.67; $Mean_{ctrl} = 74.62$, s.d. = 15.61; $p<0.05$). This finding resonates with that of Reuter *et al.* (2019) discussed previously and those of Lee and Wong (2014) suggesting that the cognitive load required to navigate a new device and learn a new concept could have contributed to the poorer performance of the AR group. This is also the case in the study of Guo and Kim (2020) which found statistically significant higher mental workloads in the AR group compared to the control group based on responses to a questionnaire. Notwithstanding that the sample sizes of these studies are insufficient to generalise the reported outcomes, these contradictory findings highlight some of the challenges with introducing novel technologies to classrooms.

2.9 Implications of Findings

A key observation from this chapter is the use of bespoke engineering-relevant DGs, AR and VR applications in the reported studies. Whereas DEGs have been reportedly used in other domains for teaching and learning, it is often rare to find their applications in engineering education as this review shows. The implication of this is that the use and adoption of these technologies would be influenced by the availability of engineering-relevant applications. Designing educational DGs is a nontrivial task that requires game creation skills and subject knowledge as Udeozor *et al.* (2022) pointed out. A possible solution could be promoting cross-disciplinary projects that would require collaborations between engineering students and students from computer and games departments to develop educationally relevant games such as those presented in the study by Fornós *et al.* (2022).

Of particular interest to this chapter is the identification of assessment methods used by educators and researchers to measure the learning effectiveness of immersive technologies for engineering education. The findings showed the use of different assessment methods with no assessment design considerations reported. Thus highlighting the need for a structured and robust assessment framework that can be applied to any immersive learning environment. External assessment methods are still by far the most frequently used assessment methods in the field of immersive technologies for engineering education. This finding makes it obvious that the transition toward more authentic performance assessments possible with immersive learning technologies is stalled, given that similar findings a decade ago are applicable today. To facilitate a more structured and unified design of assessment of learning with immersive technologies, an assessment framework that can be easily used for both external and embedded assessment designs is required. Critical to its adoption in classrooms is the need for such an assessment framework to be effective and easy to use. In subsequent chapters, an assessment framework designed for this purpose will be presented together with its applications to all three immersive learning environments under study.

2.10 Conclusions

This chapter presents the current practices in the use of immersive technologies for engineering education. Essentially, it highlights the engineering disciplines where these are used, the games and devices used, and the experimental designs and assessment methods used to measure their effectiveness as learning tools. It also provides researchers and practitioners with handy documentation of relevant journals and conferences where relevant studies are published. The DGs and devices used for learning with AR and VR are also presented in this chapter. Furthermore, this review highlights the overwhelming preference for bespoke DGs, AR and VR relevant to engineering education. For a professional domain such as engineering, the technical nature of the courses taught is believed to influence the use of bespoke DGs which also poses some challenges to the wider adoption of immersive technologies for engineering education.

Limited by the search terms and inclusion criteria, as well as the databases and the periods considered, the outcomes of this review suggest that immersive technologies are as effective as or better than traditional teaching methods. Given the range of assessment methods and experimental designs reported, it is challenging to judge the generalisability of these outcomes. External assessment methods, particularly pre-and/or post-intervention tests and self-assessment questionnaires, were used in the majority of the studies to ascertain the learning effectiveness of these tools. Embedded assessment methods such as the use of game metrics and log data were used in only a few studies, highlighting the slow progress towards achieving authentic assessments with immersive technologies. Lastly, quasi-experimental design was used by over a third of the studies, while randomised experiment and time series designs were also popular among the papers reviewed.

2.11 Limitations

Like all reviews, the current review is limited by the search terms used, databases searched, and the periods considered. It does not claim to be exhaustive, however, it provides an overview of the uses of immersive technologies in the engineering education context based on the search terms and inclusion criteria used. Due to the focus on papers that presented an assessment of effectiveness, all the other papers that simply described the use of these technologies without measuring effectiveness were excluded from consideration. This means that the current list of papers in this review is far from what is currently available in the literature.

2.12 Note

Part of this chapter was published in the **European Journal of Engineering Education**:

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

Chapter 3. Perceptions and Factors Affecting the Adoption of Digital Games by Engineering Students

In the previous chapter, the applications of immersive technologies in the engineering education domain were analysed to gain an understanding of current practices, particularly with regards to digital games (DGs) used, virtual and augmented reality devices adopted, and assessment methods used to measure learning. In the current chapter, the idea of DGs for engineering education is explored further. As a relatively new pedagogical tool, it is necessary to understand the perceptions of engineering students towards the use of DGs for learning and assessments, as well as to identify factors that influence the adoption of DGs by these adult learners. So far, only a few studies have investigated the perceptions of higher education (HE) students towards learning games, and even fewer with engineering student participants (*e.g.* Andreu-Andrés and García-Casas, 2011). To bridge this research gap, this chapter utilises a mixed-methods research design to identify factors that influence the adoption of digital learning games by engineering students and their overall perceptions of the use of games for engineering education. A questionnaire was developed based on the extended Unified Theory of Acceptance and Use of Technology (UTAUT 2) model. The questionnaire was adapted by rephrasing the wording to identify factors that influence the adoption of DGs by 125 engineering students in the quantitative study. For the qualitative data collection, a focus group interview was conducted with seven chemical engineering students. Results from the quantitative and qualitative studies suggest that engineering students value fun and engagement, as well as relevance to the curriculum, as factors that would influence their intentions to use DGs for engineering education. Students showed openness to the use of DGs for learning, but resistance to their use for assessment. These findings have implications for the classroom deployment of DGs. To conclude, this chapter provides insights to educators on the factors to consider during the implementation of games in HE classrooms.

3.1 Introduction

In HE, the growing need to improve the learning experiences of students and support the development of skills and knowledge required by employers and society is driving investments in innovative technologies. DGs and other immersive technologies have gained considerable attention as educational and training tools that provide realistic and experiential learning environments (Whitton, 2009). The recent rapid increase in interest in these technologies has been lately accelerated by the limitations of traditional classroom teaching and learning, and the challenges posed by the Covid-19 pandemic restrictions. Many HE institutions are investing in simulation games, as well as virtual and augmented realities for laboratory and practical skills training (Glassey and Magalhães, 2020; Udeozor *et al.*, 2022). Although there have been limited reports of the use of games in the engineering education domain compared to other domains such as health and business education, an increasing number of publications in scientific journals are now seen with different positive outcomes reported (Solmaz and Van Gerven, 2021; Gordillo, Lopez-Fernandez and Tovar, 2022; Udeozor, Russo Abegão and Glassey, 2022).

With any new technology, it is believed that understanding the perceptions of users enhances the adoption and intended outcomes of the use of that technology. Several studies have been conducted on the perceptions and views of educators, and the results have often shown that educators and educators-in-training alike, are receptive to the use of DGs and other immersive technologies for teaching (Razak, Connolly and Hainey, 2012; Beavis *et al.*, 2014; Noraddin and Kian, 2014; Stieler-Hunt and Jones, 2015; Khukalenko *et al.*, 2022). Noraddin and Kian (2014) found that 70% of the 273 surveyed university teachers had positive views of the use of DGs for HE. In the study conducted by Stieler-Hunt and Jones (2015) involving 13 educators, it was also found that educators believe that DGs are valuable teaching tools. The outcome of their in-depth, semi-structured interviews indicated that most teachers adopt games for classroom use out of a personal conviction about their efficacy and this is often complemented by the professional development training received. A large-scale survey of over 20000 educators regarding their attitudes towards virtual reality (VR) games for education showed moderately positive views on the subject (Khukalenko *et al.*, 2022). Although 80% of the educators believed that VR improves the academic achievements of students, 90% of the participants highlighted the need for adequate training for successful classroom integration of the technology. Generally, the outcomes of studies have shown that educators have positive views of the use of games for teaching and that many educators are currently using or are open to using these in the classrooms. While the views of educators towards the use of innovative learning technologies are almost always positive, this may differ for students. So far, some studies have investigated the perceptions of students and factors affecting their adoption of DGs for learning (Pando-Garcia *et al.*, 2016; Saleh *et al.*, 2014; Ramírez-Correa *et al.*, 2019). However, there are limited studies exploring perceptions towards educational DGs from the viewpoint of engineering students. This chapter, therefore, presents an evaluation of the perceptions of engineering students towards the use of educational DGs and the factors that affect their adoption of DGs for education.

The use of DGs for engineering education has been found to offer similar, and sometimes better outcomes when compared to conventional teaching methods (Udeozor *et al.*, 2022). As discussed in Chapter 2, DGs reportedly led to comparatively better outcomes in the test scores of students who learned with them than in those who used conventional learning methods (*e.g.* Suescún *et al.*, 2018; Xenos and Velli, 2020). Notwithstanding the positive outcomes seen, the adoption and the use of DGs for engineering education are limited when compared to domains like medical and business education (Con*nolly et al.*, 2012; Boyle *et al.*, 2016). As a professional discipline requiring certain professional competencies that are often contextual to the work environment, engineering education can benefit from the use of DGs to complement current educational practices. DGs have proven to be effective in simulating realistic environments and enhancing motivation and learning performance. When applied to engineering education, DGs could enhance the development of professional skills of graduates and contribute to educating job-ready employees for the chemical industry.

To enhance the adoption of DGs, or any new technology, and ensure that intended outcomes are achieved, it is believed that understanding the perceptions of students is of great importance (Beavis *et al.*, 2015). A few studies have explored the views of students toward game-based learning and the outcomes of most of these have been generally positive (Thanasi-Boçe, 2020; Yue & Tze, 2015; Sevim-Cirak & Yıldırım, 2020; Andreu-Andrés & García-Casas, 2011). In the study of Bolliger *et al.* (2015), 81% of the 222 participants agreed that games offer opportunities to experiment with knowledge. 58% of the 51 participants in another study agreed or strongly agreed with statements measuring their views on the efficacy of DGs to improve learning (Yue & Tze, 2015). Most recently, the outcomes of another study showed that engineering students have positive sentiments toward using VR games for education (Udeozor *et al.*, 2021). Although these studies show that students are open to the idea of using games for learning, fewer studies have explored the factors that affect their adoption of DGs for education. Using conceptual models, such as the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT), some studies identified “perceived ease of use” and “perceived usefulness” (Fagan *et al.*, 2012; Bourgonjon *et al.*, 2010) and “enjoyment” (Beavis *et al.*, 2015) as factors influencing the adoption of games for learning. However, the majority of these studies involved primary and secondary school students whose outlook toward the subject of game-based learning may differ from those of HE students, and importantly, students in engineering disciplines.

Understanding the perceptions of students and factors that influence their adoption of DGs for engineering education is crucial to informing the design and deployment of effective games. It will also enable educators to utilise appropriate instructional designs based on the needs of students when using DGs for engineering education. This chapter utilises a mixed-methods research design to identify factors that would influence the use of DGs by engineering students and to evaluate their perceptions of game-based learning for engineering education. It builds upon the outcomes of a previous study that explored the perceptions of engineering students and professionals towards the use of VR games for learning (Udeozor *et al.*, 2021). The current chapter provides solid evidence of the views of engineering students with a larger sample size that allows for a more sophisticated statistical analysis that can identify factors that have the strongest influence on the use of educational DGs. Following the sequential explanatory design (Creswell, 2011), quantitative and qualitative data are used to provide a holistic understanding of the opinions of engineering students about learning with DGs. So far, to the best knowledge of the author, only a few studies have examined the perceptions of engineering students from a quantitative viewpoint (*e.g.* Andreu-Andrés and García-Casas, 2011). Additionally, concerns have been raised about the effectiveness of existing technology acceptance models commonly adopted for games-related studies (Hsu & Lu, 2004). The current study, therefore, begins to answer the question of what factors affect the use of DGs by engineering students, by using both qualitative and quantitative data. Hence, the goal of this study is to identify factors that affect the intentions to use DG for engineering education. This study also explores the perceptions of students towards the use of DGs for learning assessment. The results of this mixed-methods study will help inform the implementation of engineering education-relevant DGs and more appropriate technology acceptance models relevant to game-based learning.

3.2 Theoretical Framework and Hypotheses Development

To identify factors that influence the adoption of DGs for engineering education, the extended Unified Theory of Acceptance and Use of Technology (UTAUT 2) model was used. The UTAUT 2 model (Venkatesh *et al.*, 2012) is an extended version of the UTAUT model.

The original UTAUT model was developed by Venkatesh *et al.* (2003) taking into account eight prominent models and theories: the Technology Acceptance Model (TAM), Theory of Planned Behaviour (TPB), Theory of Reasoned Action, Motivational Model, Model of PC Utilisation, Innovation Diffusion Theory, the Combined TAM and TPB, and Social Cognitive theory. In 2012, Venkatesh and colleagues extended the UTAUT by incorporating three new constructs: hedonic motivation, price value and habit, to form the UTAUT 2 model (Venkatesh *et al.*, 2012). Compared to UTAUT, UTAUT 2 is reportedly able to explain up to 74% (56% for UTAUT) of variance in the behavioural intentions to use new technology, and 52% (40% for UTAUT) of the variance in technology use behaviour (Venkatesh *et al.*, 2016; Wang *et al.*, 2020). The UTAUT 2 model was therefore chosen for this study as it offers better predictive power and has been widely used in related research (see Kumar and Bervell, 2019; Ramírez-Correa *et al.*, 2019; Samsudeen and Mohamed, 2019; Toyoda *et al.*, 2020; Wang, Wang and Jian, 2020).

The UTAUT 2 theorises that Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), Facilitating Conditions (FCs), Hedonic Motivation (HM), Price Value (PV) and Habit (H) are direct determinants of an individual's Behavioural Intentions (BIs) to use technology, while FCs, HM, PV and H are direct determinants of Use Behaviour (USE). These relationships are moderated by age, gender, and experiences, as shown in figure 3.1.

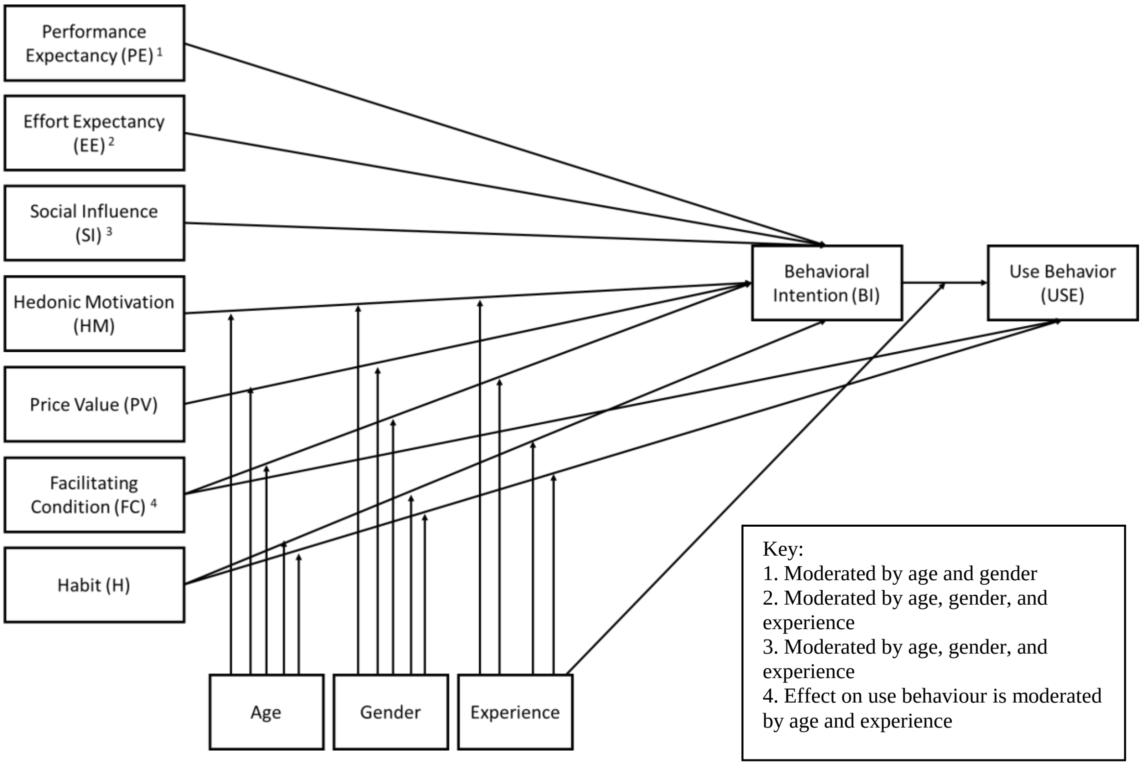


Figure 3.1: The extended Unified Theory of Acceptance and Use of Technology (UTAUT 2) model (adopted from Venkatesh, Thong and Xu, 2012).

In this chapter, a modified version of the UTAUT 2 model, shown in figure 3.2, is adopted to determine the factors that influence the BIs of engineering students to use digital learning games. USE is not considered in this study as students have not yet used DGs for classroom learning, hence intentions to use rather than use behaviours are examined. PV and H were excluded from the model because these are considered to have an insignificant impact on the student population under study. PV has to do with the monetary costs that consumers of a given technology would have to bear and how this might affect their use of such technology (Venkatesh *et al.*, 2012). In the case of the engineering students under study, they will unlikely be required to purchase game-based learning resources, since relevant learning resources are often provided to them. Hence, the cost associated with DGs is unlikely to be an influential factor. Habit is operationalised as prior use. For this study, it can be viewed as the effect of prior use of DGs on the intentions of students to use DGs for engineering education. It has been found that the effect of H is more pronounced on USE than on BIs (Limayem *et al.*, 2007). Since this study is only aimed at predicting BIs but not USE, and the fact that DGs have not been previously used for education by the participants of this study, the H construct is excluded from the model.

The effect of age, gender and experience on the relationships between the measured constructs are not considered relevant to the current study due to the relative homogeneity in the age and experience of the population under study. The ages and digital gameplay experiences of the student participants are expected to be similar as was found in a previous study (Udeozor *et al.*, 2021). In that study, out of the 27 engineering students from three different cohorts surveyed, 72% were between ages 18 and 20 years, and 100% of the students reported prior digital gameplay activities. Given that similar cohorts of students are under study, no age and experiences-based differences are expected. In engineering disciplines, it is common to find more male than female students. This was the case in the preceding study where the participants were made up of 67% male and 33% female students (Udeoz*or et al.*, 2021). Moreover, the results of prior research on gender differences in user experience with immersive technologies have been mixed. Some studies found gender differences in favour of men (Sagnier *et al.*, 2020; Wang & Wang, 2008), while others favoured women (Dirin *et al.*, 2019). Nonetheless, it is well known that there are inherent gender-based differences in technology adoption and use (Goswami & Dutta, 2016). However, due to the relatively small sample size of this study with significantly more male than female students, gender-based differences were not taken into account. Given the proportion of male-to-female participants, the data will be skewed, and the results of any analysis that takes gender into account will likely be misleading. For these reasons, age, experience and gender moderators are not taken into account in this study.

3.2.1 *The modified UTAUT 2 model*

As earlier described, a modified version of the extended Unified Theory of Acceptance and Use of Technology (UTAUT2) is used in this chapter. The six constructs of interest and their interrelationships presented in figure 3.2 are explained below. Five hypotheses to test in the study are also proposed.

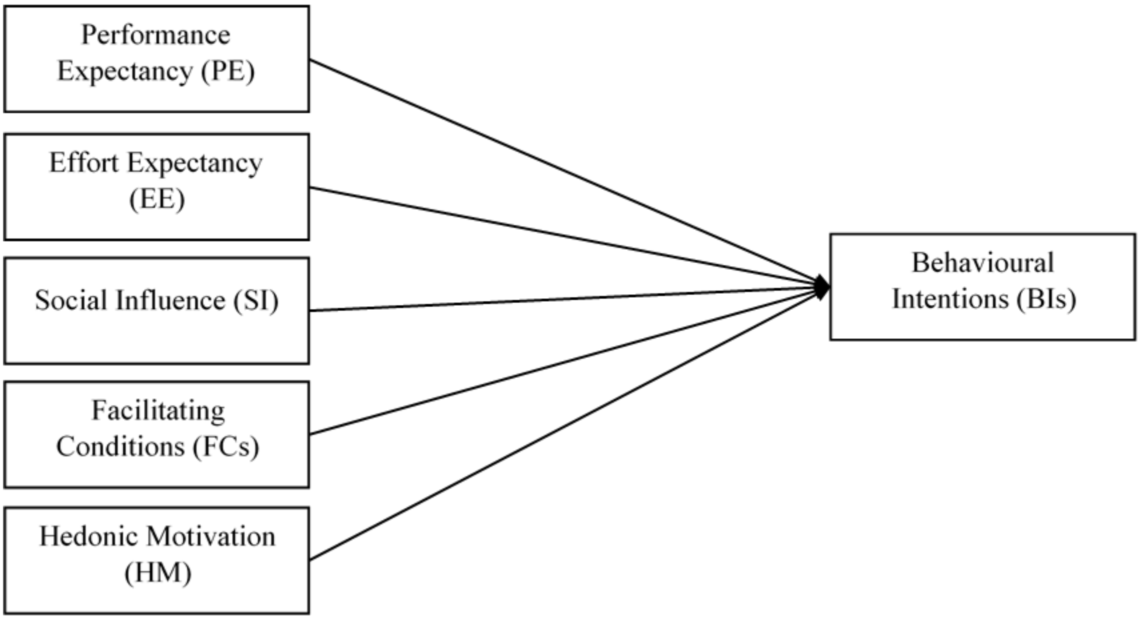


Figure 3.2**:** The modified UTAUT2 model adopted in this study to measure the factors that significantly influence the intentions of engineering students to use DGs for education.

PE is defined as the degree to which an individual believes that using a given technology will enable them to perform a certain activity. This variable represents the usefulness of the usage of a given technology, and it is often considered the strongest predictor of the BIs to use a new technology (Venkatesh *et al.*, 2003). DGs can improve the learning experiences of students, making them learn better and perform better when assessed, hence it is expected that hypothesis H1 be formulated as:

H1: *PE has a significant influence on the BIs to use DGs for learning*.

EE is defined as the perception of the degree of ease associated with the use of a given technology. Like PE, EE is often found to be a significant determinant of BIs (Venkatesh *et al.*, 2003; Fagan, Kilmon and Pandey, 2012). Therefore, it is hypothesised that:

H2: *EE has a significant influence on the BIs to use DGs for learning.*

SI is defined as the extent to which an individual perceives that people who are important to them believe that they should use a given technology (Venkatesh *et al.*, 2003). Regarding the use of DGs for engineering education, SI can be viewed as the effect of the social environment, for instance, teachers, peers, families and friends on the BIs of students to use digital learning games. Hence, it is predicted that:

H3*: SI has a significant influence on the BIs to use DGs for learning*.

FCs refer to the extent to which one believes that appropriate infrastructure exists to enhance the use of a given technology (Venkatesh *et al.*, 2003). These can be viewed as elements or factors that enhance or hinder the adoption of a given technology. In the case of learning games, this could include the support, training or skills needed to navigate the game. It is therefore expected that:

H4: *FCs have a significant influence on the BIs to use DGs for learning.*

HM is considered as the fun or pleasure one derives from using a given technology (Venkatesh *et al.*, 2012). HM is conceptualised as perceived enjoyment, and has been found to have the greatest influence on intentions to play games (Ha *et al.*, 2007). DGs are intrinsically motivating because of the elements of fun and engagement associated with them. It is therefore expected that this factor will play a major role in the intentions of students to use digital learning games. Hence, from hypothesis H5 it follows that:

H5: *HM has a significant influence on the BIs to use DGs for learning*.

BIs describe the intentions, likelihood or probability that an individual will use a new technology (Venkatesh *et al.*, 2012). BIs play a major role in determining the adoption and actual use of a new technology. Nevertheless, this chapter does not attempt to predict nor explain the actual use of DGs by students but aims at explaining factors that affect their BIs to use DGs for engineering education.

3.3 Research Methodology

A mixed-methods research design was used to collect and analyse quantitative and qualitative data for this study. A mixed-methods design was adopted for this study because no existing validated questionnaire that addresses the goals of this research was found. Therefore in addition to the quantitative data collection using the modified UTAUT2 questionnaire that examines influential factors for the acceptance of DGs by engineering students, qualitative data was collected on the perceptions of students toward the use of DGs for learning assessment. Mixed-methods research design applied herein is considered an advanced methods procedure as it requires an understanding of both quantitative and qualitative research (Creswell, 2011). It is used to gain a better understanding of research problems and questions that either quantitative or qualitative methods alone cannot provide (Creswell, 2011). The mixed-methods research design used for this study follows Creswell's explanatory sequential design, which involves the collection of quantitative data, followed by qualitative data, as shown in figure 3.3. This research design allows for an in-depth exploration of findings from quantitative data through qualitative data collection. It also provides an opportunity for the elaboration of opinions that may not have been possible with only quantitative data collection. For this study, an online questionnaire and a focus group interview were used for quantitative and qualitative data collection, respectively. Methods triangulation strategy which involves the collection and comparison of data from multiple methods is used to validate the interpretations of the qualitative data collected (Carter *et al.*, 2014).

Figure 3.3: Explanatory Sequential Design (adapted from Creswell, 2011).

3.4 Quantitative study

3.4.1 Participants

The sample size for this study is 125. These participants were chemical engineering students from three European universities - Newcastle University, UK, Imperial College London, UK, and KU Leuven, Belgium, who volunteered to take part in the study. A convenience sampling method was used to recruit these participants (Creswell, 2011). As shown in table 3.1, the participants were made up of 70% male and 30% female students with 99% of the students below 29 years of age.

3.4.2 Questionnaire design

The goal of the quantitative phase of this study is to identify factors that affect the adoption of educational DGs by engineering students. To achieve this, an online questionnaire was utilised. As presented in table 3.1, the questionnaire collected demographic data of participants: age, gender, gameplay experiences and finally their perceptions of game-based learning measured on 6 constructs based on the UTAUT 2 model. Although demographic information of the participants was not eventually used in the model analysis, they were collected to see whether an equal proportion of male and female responders took part in the study. This would have allowed for the inclusion of gender as a moderator in the modified model used in this study. The wording of each question was modified to suit the study context, in this case, the statements were modified to ask participants their opinions on game use rather than on “the system” as in the original UTAUT 2 questionnaire. For instance, the original UTAUT 2 item for PE 1 is “I would find the system useful in my job”, whereas, for this study, PE 1 is rephrased thus: “I think that I would find digital games useful for learning engineering modules” to match the context of the study. The original UTAUT 2 questionnaire can be found in the paper by (Venkatesh et al., 2012)The constructs were measured on a 6-point Likert scale, with 1 used for “Strongly Disagree” and 6 for “Strongly Agree”.

Table 3.1: Gameplay experience data of (n=125) surveyed participants.

		Absolute-frequency	**Percentage**
Prior gaming experience	Yes	117	93.6
	No	8	6.4
Gaming habits	Less than 5 hours	54	43.2
	5-10 hours	41	32.8
	11-20 hours	17	13.6
	21-30 hours	7	5.6
	31 hours or more	1	0.8
	Unspecified	5	4.0
Game enjoyment	Low (1-4)	14	11.2
	Average (5-7)	49	39.2
	High (8-10)	58	46.4
	Unspecified	4	3.2

A 6-point Likert scale was chosen as it is believed to produce a higher discrimination and reliability trend than a 5-point scale (Chomeya, 2010). Before using the questionnaire for data collection, validity checks were carried out by academic experts. They checked the grammar, content and length of the questionnaire to ensure the validity of the modified UTAUT 2 questionnaire for the study purpose. Pilot testing of the questionnaire was done by volunteer postgraduate chemical engineering students at Newcastle University. This group of students completed the questionnaire, highlighting items that were difficult to understand. Based on feedback, changes were made to the wording of the questions where needed.

3.4.3 Data collection and analysis

Data were collected from the student participants on three different occasions. For all three data collection events, students also took part in gameplay sessions. First, participants were requested to complete the online questionnaire to evaluate their perceptions of DGs for engineering education and their gameplay experiences. Participants filled out the questionnaire before proceeding to the gameplay sessions. CosmiClean game, a recycling game described in Chapter 4 was planned by students. For the current chapter, no game data was used for the analysis and all the data used here were collected prior to gameplay, hence, could not have influenced the outcome of the analysis. Before utilising this questionnaire for data collection, ethical approval was obtained from the Faculty of Science, Agriculture and Engineering ethics committee at Newcastle University permitting the study to be carried out. The age of the target population, the type of data to be collected and how these data were intended to be used were fully documented in the application form. For the data collected throughout this research project, the security of data was ensured by storing them in the university server and for the duration of the project. No personal information such as full name, date of birth and address of the participants that could be linked back to them was collected.

The multivariate analysis known as structural equation modelling (SEM), was used for the analysis of the quantitative data. SEM is useful for analysing models as it allows for the simultaneous analysis of relationships between constructs in a given model (Hair *et al.*, 2017). The two types of SEM analysis used in research are covariance-based SEM (CB-SEM) and partial least square SEM (PLS-SEM). The CB-SEM is generally used to test and confirm theories, while the PLS-SEM is used in exploratory studies to develop or extend an existing structural theory (Hair *et al.*, 2017). Using the modified UTAUT2 theoretical model, this study finds it appropriate to use the exploratory PLS-SEM to determine the predictors of the BIs of engineering students towards DGs for education. It is worth emphasising that the modification made to the original UTAUT2 questionnaire items presented in table 3.2 involved rephrasing the items to match the context of the study, as mentioned above. The

PLS-SEM measurement model and structural model analyses for this study were conducted (Hair *et al.*, 2017) using the SmartPLS^{TM} Version 3 software.

To determine the minimum sample size for these analyses, Cohen's statistical power analysis for multiple regression models was carried out (Cohen, 1992; Nitzl, 2016). This minimum sample size estimation method is considered more rigorous and appropriate for determining the minimum sample size for PLS-SEM analysis (Kock and Hadaya, 2018; Hair *et al.*, 2021). Based on this approach, the minimum sample size for this analysis was found to be 92, based on 5 independent variables and working with a statistical power of 0.8, a significance level of 0.05 and a medium effect size (Cohen, 1992; Nitzl, 2016). Therefore, a sample size of 125 used in this study is considered sufficient for the current analysis.

3.5 Results of the Quantitative Study

3.5.1 Measurement model evaluation

The measurement model aims to evaluate the reliability and validity of the constructs of the model. Cronbach's alpha (α) and composite reliability (*CR*) were calculated to measure the internal consistency reliability of the constructs. These were calculated automatically by the SMART-PLS software, but the following equations 3.1 and 3.2 can be used to calculate *CR* and α, respectively:

$$CR = \frac{(\sum_{k=1}^{K} l_k)^2}{(\sum_{k=1}^{K} l_k)^2 + \sum_{k=1}^{K} var(e_k)} \qquad \textit{Equation } 3.1$$

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

$$\alpha = \frac{K \cdot \bar{r}}{[1 + (K-1) \cdot \bar{r}]} \qquad \textit{Equation } 3.2$$

where: *K* = 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)

Table 3.2: List of measurement items in the online questionnaire used to measure the perceptions of engineering students towards the use of DGs for education.

Measurement constructs	**Item codes**	**Items**
Performance Expectancy	PE_1	I think that I would find digital games useful for learning engineering modules.
	PE_2	I think that playing digital games designed to teach engineering concepts would increase my engineering knowledge.
	PE_3	I think that playing educational digital games would enable me to learn more quickly.
	PE_4	Using digital games to learn engineering modules will increase the quality of my learning experience.
Effort Expectancy	EE_1	I expect that my interaction with a well-designed engineering digital game would be clear and understandable.
	EE_2	I expect that it would be easy for me to become skilful at the game.
	EE_3	I expect to find games designed to teach engineering concepts easy to play.
	EE_4	I expect that operating such games would be easy for me.
Social Influence	SI_1	My teachers would expect me to use digital games for learning if made available.
	SI_2	My peers will be supportive of my use of digital games for learning.
	SI_3	My teachers will be very supportive of my use of digital games for learning.
Facilitating Conditions	FCs_1	I have the resources necessary to play digital games for learning purposes.
	FCs_2	My university will provide the necessary support for me to use digital games for learning.
	FCs_3	Playing digital games for learning fits well with the way I learn.
Hedonic Motivation	HM_1	I will really enjoy playing games to learn.
	HM_2	I think that playing engineering games will be fun.
	HM_3	I think that playing engineering games will be very entertaining.
Behavioural Intentions	BIs_1	After the Recycling game, I intend to use digital games to learn again in the near future.
	BIs_2	I will continue to use digital games to learn engineering principles if made available to me.
	BIs_3	I am open to using digital games to improve my knowledge of chemical engineering principles if made available to me.

Table 3.3: Reliability and validity measures of the modified UTAUT2 model used to measure the perceptions of (n=125) engineering students toward DGs for education.

Constructs	**Item codes**	**Mean score**	**Standard deviation**	**Factor loading**	***AVE***^{a}	$\boldsymbol{\alpha}$	***CR***^{a}
Performance Expectancy (PE)	PE_1	4.36	0.90	0.864	0.681	0.847	0.895
	PE_2	4.67	0.84	0.794			
	PE_3	4.50	0.90	0.802			
	PE_4	4.53	0.96	0.839			
Effort Expectancy (EE)	*EE_1				0.764	0.845	0.906
	EE_2	4.62	0.94	0.81			
	EE_3	4.51	0.91	0.896			
	EE_4	4.67	0.95	0.913			
Social Influence (SI)	SI_1	4.63	1.03	0.724	0.674	0.784	0.86
	SI_2	4.30	1.03	0.892			
	SI_3	4.76	0.95	0.838			
Facilitating Conditions (FCs)	*FCs_1				0.757	0.707	0.861
	FCs_2	4.52	1.14	0.785			
	FCs_3	3.97	1.31	0.948			
Hedonic Motivation (HM)	HM_1	4.02	0.99	0.936	0.878	0.931	0.956
	HM_2	4.18	0.94	0.947			
	HM_3	3.96	1.07	0.928			
Behavioural Intentions (BIs)	BIs_1	3.40	1.15	0.888	0.753	0.836	0.901
	BIs_2	3.64	1.17	0.894			
	BIs_3	4.26	1.22	0.819			

*Removed due to low outer loading (<0.7); ^{a}Average Variance Extracted; ^{b}Composite Reliability; Item codes represent individual questions associated with given constructs.

As shown in table 3.3, the α and CR scores were all above the 0.6 minimum scores (Hair *et al.*, 2017). This shows that the measurement constructs have strong internal consistency reliability, indicating good correlations between items intended to measure the same constructs. Furthermore, the validity of the constructs was determined by evaluating the convergent and discriminant validities as recommended by Hair *et al.* (2017). The factor loadings for each survey question/item and the average variance extracted (AVE) for each construct (Equation 3.3), shown in table 3.3, were greater than the recommended minimum values of 0.708 and 0.5, respectively.

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

Additionally, the Heterotrait-Monotrait ratio (*HTMT*), a measure of discriminant validity, shown in table 3.4, was found to be below the 0.9 recommended maximum value (Gold *et al.*, 2001). These values indicate that the discriminant and convergent validity conditions of our model were met. Hence the constructs and items can be considered both valid and reliable.

Table 3.4: Heterotrait-Monotrait ratio (*HTMT*) measure of discriminant validity.

	BIs	**EE**	**FCs**	**HM**	**PE**	**SI**
Bis						
EE	0.357					
FCs	0.548	0.352				
HM	0.764	0.331	0.493			
PE	0.574	0.360	0.781	0.577		
SI	0.413	0.494	0.408	0.451	0.593	

3.5.2 Structural model evaluation

Having established the validity and reliability of the constructs with the measurement model, the next step was the evaluation of the structural model. This involved ascertaining the coefficient of determination (R^2) and the significance of path coefficients (β) (Ha*ir et al.*, 2017). Before proceeding with these analyses, the model was evaluated to ensure there were no collinearity issues (Hair *et al.*, 2017). It was found that all constructs had variance inflation factor (*VIF*) values of less than 3, which is the recommended threshold, indicating that there are no multi-collinearity issues as evident from table 3.5.

Table 3.5: Results of full collinearity test for each measured construct in the modified UTAUT 2 model.

Constructs	**Variance inflation factors**
EE	0.357
FCs	0.548
HM	0.764
PE	0.574
SI	0.413

To evaluate the structural model, the path significance of the relationships between constructs was determined using a (BCa) bootstrapping sampling technique on 5000 sub-samples following the recommendations of Hair *et al.* (2017). Bootstrapping is a resampling technique used to correct sampling biases in data. It does so by randomly resampling the data multiple times with replacement and then recalculating the bootstrap statistics based on the samples. In PLS-SEM, BCa bootstrapping is useful for overcoming the effects of multivariate non-normality data (Green, 2016). Figure 3.4 shows the relationships between the constructs, the path coefficients and their influence on the model.

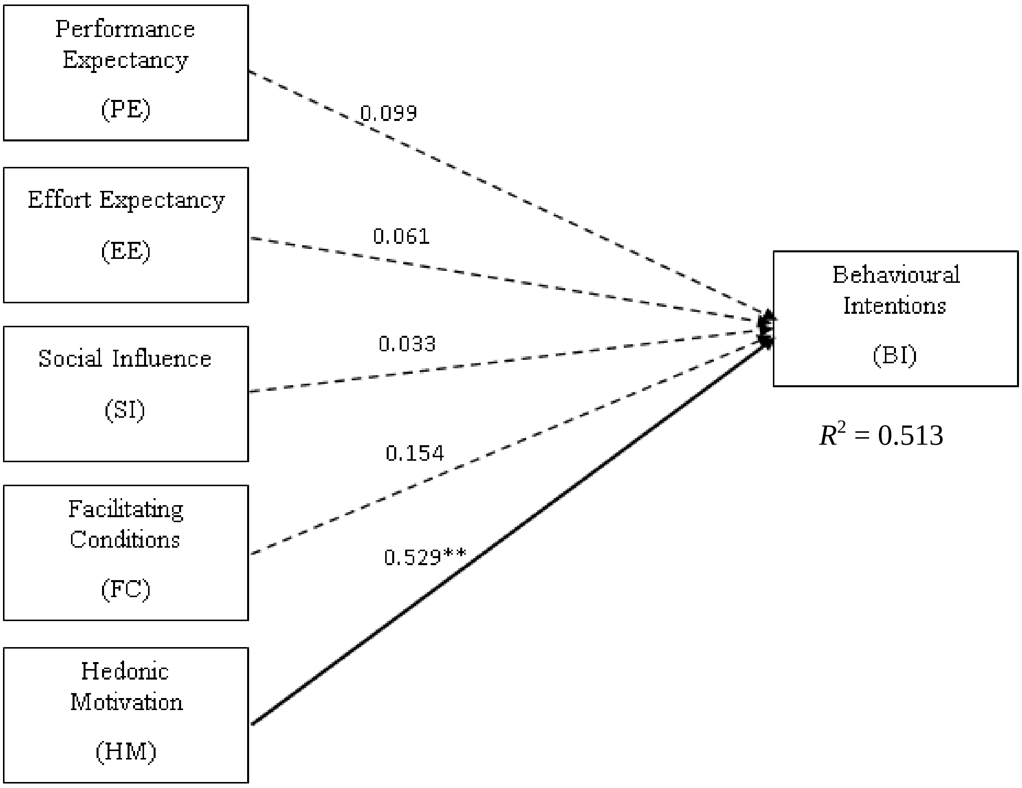


Figure 3.4: Outcome of a structural equation modelling analysis to predict factors that influence behavioural intentions of engineering students towards DGs for education (**$p < 0.001$).

From figure 3.4, the variances explained by PE, EE, SI, FCs and HM for the BIs of students to adopt DGs for learning are 0.099, 0.061, 0.033, 0.154, and 0.529, correspondingly. From these results, it was found that only HM had a significant positive influence on the BIs of students, supporting H5. Therefore, hypotheses H1, H2, H3, and H4, which assumed that PE, EE, SI, and FCs, correspondingly, have a significant influence on BIs, are rejected. Given that $R^2 > 0.2$, the modified UTAUT 2 model is considered acceptable for this behavioural study (Hair *et al.*, 2017), as it is able to accurately predict more than the recommended 20% (51.3% in this case) variance in the BIs of engineering students to use DGs for learning.

3.6 Qualitative Study

The goal of this second phase of the current study is to understand the perceptions of students towards digital learning games for engineering education. It also explores further the factors that affect their use of DGs by eliciting additional information that might not have been captured in the survey. In the quantitative study, HM was found to be the only influential factor in the BIs of students to use games for engineering education. In this qualitative phase, the reason for this is explored as well as other factors that students consider potentially relevant that were not captured in the quantitative study. Furthermore, the views of students on the use of games for assessment are explored. This aspect was not addressed in the quantitative study due to the use of a pre-existing technology acceptance model. The questions asked during the interview centred on three main areas:

RQ1: What are the views of students on the use of DGs for engineering education?

RQ2: What attributes do students expect from a DG for engineering education?

RQ2: What are the views of students towards DGs for assessments?

3.6.1 Participants

Seven volunteers who took part in the quantitative study participated in the focus group interview. The participants were chemical engineering students from Imperial College London. These were the cohort of students that were accessible at the time of this study due to the restrictions posed by COVID-19 pandemic. There were four female and three male student participants.

3.6.2 Data collection and analysis

A one-hour focus group interview was conducted with all seven participants to further explore the quantitative data collected in the first study. A semi-structured interview approach was used (Creswell, 2011). Due to the Covid-19 pandemic restrictions, the interview took place remotely via Microsoft Teams. Permission was obtained from the participants to record the discussions. Appropriate ethical approval from the ethical board at Newcastle University was also received for this study. Both video recordings and transcripts were collected during the discussions. Descriptive coding method also known as Topic coding was used for this analysis (Saldana, 2009). The analysis of the qualitative data followed a systematic approach. First, the automatically generated transcripts were carefully read through, compared against the video recording and edited where necessary for accuracy. Next, responses to each question were extracted. To identify the main points from the discussions, the responses to each question were read and key points (codes) were highlighted. For RQ2, the key points were compared to the measured constructs of the UTAUT 2 model used in the quantitative study, and the links to the constructs were noted. For RQ1 and RQ3, recurring points from the responses of the participants to the interview questions were noted as well as the frequency of appearance. This process was manually carried out in Microsoft Word because of the low number of participants involved. With key codes highlighted, points that convey similar ideas were grouped to form themes presented in table 3.6. The coding process and results were evaluated by two academics to ensure the reliability of outcomes.

3.7 Results of the Qualitative Study

RQ1: Views of students on the use of DGs for engineering education

Theme 1: Positive views

When asked if DGs could be used for engineering education, five students (71%) agreed to some degree that DGs are suitable for engineering education. Most of the participants were of the opinion that games can be used to learn, but for engineering education, the opportunities were thought to be limited to some modules:

> 'I think you could learn something entirely through games…'
>
> 'I reckon, something like heat transfer could be learned entirely through gaming… …visualise a lot better.'
>
> '… I think for things like projects (would) work with games... so more [useful] for project work [knowledge application] versus learning content [knowledge acquisition].'

Theme 2: Negative views

Two participants did not think that games were suitable for engineering education. They believed that games could not be useful for every engineering subject and that it would be practically difficult to replicate engineering systems in games:

> '…there's nothing that you can learn through a game that you cannot learn by just reading some [pages] in a short amount of time…'
>
> '…because of how complex real systems [in engineering] are, it's often really hard to mimic those in a game setting.'

Additionally, all seven participants expressed the view that games should be used as supplements or add-ons rather than as stand-alone pedagogical tools. These responses show that, although the majority of the students perceive games to be useful for engineering education, there are some concerns about their suitability and practical use.

Table 3.6: Matrix for coded interview responses from (n=7) engineering student participants.

Research questions	**Number of participants**	**Codes**	**Themes**
Views towards games for engineering education	5	Effective for learning Enhance visualisation Good for skills learning Experiential learning Project-based learning	Positive views
	2	Not better than existing pedagogies Faster to learn the existing way	Negative views
Qualities expected of engineering educational games	3	Collaboration Competition Leaderboard	Motivation/engagement
	3	Clear objectives Progressive learning	Relevance
	3	Visual/aesthetically pleasing More than just for learning	Quality design
Views on games for assessment	7	Changes exam revision strategies Extra stress	Test anxiety
		Poor gaming skills might affect grades	Gaming skills interference
		Glitches and poor connectivity	Technical concerns
	1	Depends on how it is introduced	Implementation practicalities

RQ2: Qualities expected of DGs for engineering education

Participants were also asked to describe qualities that would positively influence their adoption and use of DGs. Comparing the results of the qualitative data to the quantitative study, the themes that emerged from the responses are:

Theme 1: Motivation/engagement

This was the most frequently occurring theme in the data. This aligns with the observations in the quantitative phase of this chapter, in which HM was also found to be the most significant factor for DG use. Participants mentioned some key factors that would enhance their gameplay experience and influence their adoption of games for learning. Features of collaboration and competition were mentioned by these students as qualities they would appreciate in educational games:

> '…Say that aspect of competitiveness is always helpful in sense…, being able to play against other people.
>
> Always helps to like make it more interesting, because then it's not just us, [it's] you and your friends, seeing how you will do in comparison to others…'
>
> '…Important playing with someone else makes a game infinitely more fun…'

Theme 2: Relevance

Participants expressed the need for any DG introduced in the classroom to be relevant to their learning journey. They expect to see a clear link between learning objectives and the proposed game objectives. Relevance here is considered as the perceived usefulness, which in the UTAUT 2 model is conceptualised as PE. Here are what some of the students had to say:

> '…So I think you should make it clear what exactly you're teaching and not [just an] educational game because if someone's like, trying to revise through a game…'
>
> '…you [should] learn something new [with the game]…'
>
> '…I feel like the really good games have an objective like Mario international game you saved princess. You know, like, really clear cut…
> …for an educational game, [it should say] 'you need to do this at the end of the time…
> your outcomes will be, you will have completed the learning objective x y …''

Theme 3: Quality game design

The last theme that emerged from the data was "quality game design". Participants mentioned the need for an educational game to be aesthetically pleasing and do more than just teach. For these students, games should be well designed, and games should do more than just replicate what students can learn from existing pedagogical tools:

> '…like the visualisation will have to be good, done by someone who knows some software…, it'll have to be together and feel like a complete kind of game.'
>
> '… [What] I'd like to say is that even educational games, setting education as the objective, might be a tad boring…

> …goal itself should be different because if the end goal is education, why am I playing a game? I could just read that [from books] to also achieve that end goal.’

Another student elaborated on the need for a good educational game to provide instant and continuous feedback. This student also mentioned the need for a leaderboard that would tell players how well they are doing against pre-set goals.

Having heard the views of the students regarding the qualities they expect from an educational game for engineering education, students were subsequently asked to talk about engineering modules/courses that would benefit from game-based learning. Modules that generally require students to operate or design processes, and/or products, were described as those that would most likely benefit from DGs. As the participants of this study were chemical engineering students, the following chemical engineering modules were mentioned: Fluid Mechanics, Heat and Mass Transfer, Safety, Separation Processes, Reaction Engineering, and Design Processes.

RQ3: Views of students towards DGs for assessment

Participants were asked if they would be happy to be assessed with DGs. Unanimously, they said “no”. When probed to give reasons for this, the students gave several reasons which are described in the themes below.

Theme 1: Test anxiety

The common point raised by the participants was that assessment with DGs would create extra layers of stress and anxiety. They reiterated that traditional assessments already cause some levels of stress to students and that changing the format of the assessments would not make things any easier. Additionally, it was said that changing to game-based assessment would require changing the way students study and prepare for exams:

> ‘...I wouldn't, I would never really want to have games replace my exams, just because I'm used to like revising a certain way for an exam and knowing kind of what to expect. So, even now, like you saw how like online exams. People were panicking and that's not even changing…’
>
> ‘…But yeah, I think just having the idea of a game as an assessment, adds a level of extra stress that we don't really like. I love games, but I still wouldn't want to do a game as an assessment…’

Theme 2: Gaming skills interference

In addition to the perceived stress students anticipate with game-based assessment, the effect of gaming skills on performance was also raised. Participants worried that performance in educational DGs would be influenced by prior gaming experiences and the gaming skills of the students:

> '…but I think the [game-based] examination requires another skill that you have to be good at, in addition to everything you're being tested on…
>
> As I play like much less than other people and then when I do occasionally play games, I'm like way worse. It is a disadvantage…'

Theme 3: Technical concerns

One participant elaborated on the technical uncertainties associated with the use of DGs for something as important as assessment. They described the kind of technical issues that arise during gameplay and how this could adversely affect students:

> '…I still wouldn't want to do a game as an assessment because I know what can go wrong in games. I somehow broke the game we were playing twice by mistake. So just because when it comes to things like computers, especially like a lot of games. There's bugs and sometimes they're not found by even the triple-A companies upon release. So if that comes into play during an exam, if something bugs, that's an entire level of stress that I don't want to deal with…'

Theme 4: Implementation practicalities

Although in general, the participants would prefer not to be assessed with games, some of them showed some flexibility. Some students were open to the use of DGs for assessments only if these had little or no impact on their grades. One student said they would not mind being assessed with games if it accounted for no more than 5% of the module score:

> '… [I] do not want to do it as an official exam. I will enjoy it if it was not timed and if it was like worth 5% of my module…'

Another student would consider game-based assessment only if it was gradually rolled out throughout the semester and not just once at the end of the module:

'…I think I'd be open to it completely dependent on how it was implemented. So for example, if at the start of the module, the teacher was like hey I've got a new thing I want to try out this year. I want to try out again at the final exam and we're going to go through a few examples in the class of how a game-based exam might work… However, if a person just came in and slapped me down with a game and was like, plug this into your computer. This is going to be your assessment. I think I would crap myself like it's just not it…'

3.8 Discussions and Implications

This chapter aimed to identify factors that influence the BIs of engineering students towards the use of DGs for learning, as well as their perceptions of DGs for engineering education. This chapter develops further a previous study (Udeoz*or et al.*, 2021) by carrying out a PLS-SEM analysis to identify factors that influence the use of DGs by engineering students using data from a larger sample size. Using a mixed-methods research design, the results of the quantitative data analysis indicate that of all considered constructs, only HM had a statistically significant influence on the BIs of students to use DGs for engineering education. This suggests that the cohorts of students who took part in this study consider pleasure, fun and enjoyment derived from gameplay as the most important factor that would determine their use of DGs for engineering education. This finding aligns with the findings of others who examined the perceptions of the use of games by HE students (Wang *et al.*, 2020; Duffull & Peterson, 2020; Udeozor, Toyoda, *et al.*, 2021; Andreu-Andrés & García-Casas, 2011). Similar to the current finding, Wang and colleagues found HM to be the singular most significant factor that influenced the BIs of management students to use business games for learning. PE, EE and SI did not have any significant influence on the intentions of the students to use business games. Using data from a questionnaire with open-ended questions, Duffull and Peterson (2020) found that one of the emerging themes from the responses of pharmacy students regarding the use of games for learning was “fun”. They found that fun was an important factor for repeated gameplay and immersion when adopting games for learning.

Although unexpected, it is not completely surprising that only HM had a significant influence on BIs in the current study. With the majority of the cohort of students in this study being undergraduates in their early 20s, and generally regarded as digital natives (Prensky, 2001), it is understandable that they value fun, enjoyment and pleasure derived from gameplay as important factors that will determine their adoption of games for learning purposes. A contrast can be found in the study of Toyoda *et al.* (2020), which reported that for professional chemical engineers, PE had the most significant influence on their intentions to use VR for training. This difference of opinions is attributed to the value the different groups of users place on these technologies. Whereas professionals are likely to adopt immersive technologies to become better at their jobs, students on the other hand would most likely adopt these because they are fun and engaging to learn with (Udeozor, Toyoda, *et al.*, 2021). Inconsistencies in results were also found in studies investigating perceptions of HE students (Estriegana *et al.*, 2019; Malaquias *et al.*, 2018; Zulfiqar *et al.*, 2021; Pando-Garcia *et al.*, 2016). These studies found PE, EE and/or SI to be the most influential constructs on the BIs of students to use games for learning. These differences in findings might be affected by the learning styles (Wouters & van der Meulen, 2020) and potentially, the English language proficiency levels of the students under study. A learner who enjoys experimenting with things might progress through an immersive learning task faster than another learner who is most concerned with understanding the concept or logic of things. Nonetheless, studies have shown that immersive applications have similar cognitive benefits for different types of learners (Lee *et al.*, 2010). Furthermore, given that the wording of the items on each measured UTAUT 2 construct is often modified to fit research purposes, there are bound to be differences in the UTAUT 2 models which could influence the outcomes of the research. Inconsistencies and unexplainable findings in similar game-related studies using existing technology acceptance models such as UTAUT 2 have been linked to the likelihood that these models are missing some important constructs necessary for measuring perceptions and acceptance of games for learning (Hsu and Lu, 2004). The implication of this, therefore, is the need for theoretical models developed specifically to measure the acceptance of innovative learning technologies like immersive technologies in educational research.

Furthermore, the qualitative data collected and analysed in this study provided a holistic and deeper understanding of the perceptions and factors influencing the adoption of DGs by engineering students. The findings of the qualitative study showed that students are open to using games for learning and believe that DGs can be useful for engineering education. However, students emphasised the need for games to be used as add-ons rather than stand-alone pedagogical tools. There were also concerns about designing engineering DGs given the complexity of simulating engineering systems. When describing qualities that they would expect of DGs designed for learning, students mentioned the need for games to be collaborative and/or competitive, particularly in HE. They expect games to be fun and relevant to their learning, qualities that lend themselves to the HM and PE constructs, respectively. Although PE was not found to have a significant influence on the BIs of students to adopt DGs in the quantitative phase of this study, the majority of the participants clearly stated or implied that the goals of DGs used for learning should clearly align with their curriculum learning outcomes. Similar findings have also been reported in previous studies (Estriegana, Medina-Merodio and Barchino, 2019; Zulfiqar *et al.*, 2021). Lastly, the quality of DGs was mentioned as a factor that would affect the use of games for learning. Students expect educational games to be well-designed and complete to be attractive enough for learning. A well-designed game was described as one that is aesthetically pleasing with an instant and continuous feedback system. These findings highlight implications for educational DG designs and classroom implementations. As a novel pedagogical tool that is attracting a lot of attention, educators may be tempted to implement DGs without critically evaluating their quality and relevance. Introducing poorly designed DGs or worse, DGs that are not completely aligned to the curriculum may result in unintended outcomes and may risk putting learners off innovative teaching technologies in the future. There is also the need for educational DG designers to pay attention to the quality of games designed for adult learners. This may be challenging given the smaller budget sizes that educational institutions have compared to large game design companies. Nonetheless, educational DGs need to be aesthetically pleasing, fun to play and importantly, relevant to the target students.

DGs provide opportunities for authentic assessments and offer many benefits, such as the application of modern psychological theories to assessment, increasing assessment coherence and enabling the measurement of “hard-to-measure” constructs that would otherwise be difficult to achieve (Buckley *et al.*, 2021). But as a non-conventional assessment tool, it was necessary to hear the opinions of engineering students regarding the use of DGs for assessment. Although a majority of students in the focus group believe that DGs can be useful for engineering education and are willing to adopt them for learning, they did not share the same sentiments for the use of games for assessment. Similar outcomes were reported in the study by Cook-Chennault and Villanueva (2020). A few students in the current study indicated considering using DGs for assessments but on the conditions such as how they are rolled out and their weighting on the overall grades. However, most of the students believe being assessed with DGs would increase test anxiety. This is a particularly interesting finding considering that proponents of game-based assessments claim otherwise (Mavridis and Tsiatsos, 2017; Shute *et al.* 2017).

Contrary to the findings of the current study, Mavridis and Tsiatsos (2017) found that test anxiety was lower in game-based assessments compared to traditional assessments and that students had positive attitudes toward game-based assessments. This inconsistency in findings could be due to the fact that in the study of Mavridis and Tsiatsos, students already experienced assessments through games and made judgements based on that, whereas, in the current study, students merely described their perceptions of game-based assessment with no prior experience. Furthermore, students in the current study looked at game-based assessment from a high-stake assessment point of view, but in the case of the study of Mavridis and Tsiatsos, the assessment conducted was for research purposes, with little or no implication on the grades of students. The low sample size of 30 participants in their study and only seven in the qualitative study of this chapter also warrant caution in interpreting the reported outcomes. In addition to perceived test anxiety, students also mentioned gaming skills, technical issues and implementation practicalities as issues with using games for assessment. The concern that gaming skills could influence performance is quite interesting and warrants further investigation. In general, the findings of this chapter make obvious the need for more research into immersive learning pedagogy as unfounded integration of these immersive learning tools into current educational practices will likely be met with some resistance and unintended effects. DGs are non-conventional for learning, it is therefore recommended that implementing these for assessment should come after students have become familiar with learning with games. As mentioned by one of the students, rolling out game-based assessments should happen gradually and should be used for formative assessments. Doing so will likely reduce the test anxieties of students and enhance learning performance with DGs.

In summary, the elements of fun, engagement and motivation associated with HM seemed to have the strongest impact on the BIs of students to use DGs for engineering education from both the quantitative and qualitative results. PE was identified in the quantitative study as an influential construct for the adoption of DGs by engineering students. These findings from both the quantitative and qualitative phases of this study strengthen the argument for a mixed-methods research design. The additional findings from the qualitative study provided an in-depth understanding of possible factors that could play crucial roles in enhancing the adoption of DGs by engineering, and potentially, other HE students. These results could be valuable to DG designers by providing insights into factors that should be considered when designing educational games. Despite the overall positive perceptions towards games for learning, the strong and negative opinions on the use of games for assessment were attributed to reasons such as increased anxiety, changes to the old ways and gaming skills interference.

3.9 Limitations

This study benefits from the best of both quantitative and qualitative research, however, there are some limitations to the study which could inform future studies. One limitation of the quantitative study was the use of a technology acceptance model which was not specifically designed to evaluate perceptions of the use of educational immersive technologies. As no validated models exist for this purpose, future research should consider developing an immersive learning technology acceptance model appropriate for examining perceptions and identifying factors that play influential roles in the adoption of immersive technologies by learners. Furthermore, the sample size of the studies in this chapter limits the wider generalisability of the findings reported. Future studies should repeat this study using probabilistic sampling strategies (Creswell, 2011) with larger sample sizes. Participants from a range of institutions should be purposefully recruited for the studies.

3.10 Conclusions

This chapter presented an overview of the perceptions of students towards the use of DGs for engineering education using a mixed-methods research design. The outcomes of the studies showed that engineering students are open to the idea of learning with DGs but did not show the same sentiments for game-based assessments. Using a modified UTAUT2 model, it was found that only one construct had a significant influence on the intentions of students to use DGs for engineering education, whereas some other potentially important factors were identified from the qualitative study. This chapter highlighted the factors considered important for the adoption of DGs by engineering students. It also reported data suggesting that students are sceptical about the use of DGs for assessments for different resons, one of which is the concern that game experience would interfere with performance in educational DGs. This concern, however unexpected, warrants further research to determine whether good gaming skills would result in good learning performance in DGs.

3.11 Note: Part of this chapter has been submitted to the **International Journal of Educational Technology in Higher Education.** Udeozor, C., Russo Abegão, F., & Glassey, J. (submitted). Perceptions and factors affecting the adoption of digital games for engineering education: a mixed-method research. *International Journal of Educational Technology in Higher Education*.

Chapter 4. Relationship between Perceptions, Experience and Performance of Students in an Educational Digital Game

In the preceding chapter, fun and enjoyment of digital games (DGs), the hedonic motivation construct, was found to have the most significant influence on the use of DGs by engineering students. Furthermore, the use of DGs for assessment was unpopular among students, given their concerns that gameplay skills and experience might interfere with performance in an educational DG. This chapter explores this issue by evaluating the relationship between the gameplay experience of engineering students, their perceptions of the use of DGs for education and their performance in an educational DG. The belief that the perceptions and acceptance of new technologies play crucial roles in the success of such technology has led to several studies exploring the perceptions of users (*e.g.* Sevim-Cirak and Yıldırım, 2020; Thanasi-Boçe, 2020). Nonetheless, there is limited to no evidence that the perceptions and experiences of students affect their performance in games. This gap in the literature and the concerns raised by students in the previous chapter prompted the study in this chapter. This chapter employs a correlational research design to explain the relationship between prior game experience, perceptions of and performance in educational DGs. Using the extended Unified Theory of Acceptance and Use of Technology (UTAUT 2) online questionnaire, and log files from the CosmiClean gameplay of students, Spearman's correlational analysis was carried out (Cohen *et al.*, 2007). The findings suggest that no relationship exists between the perceptions of students towards games for learning and their gameplay performance. However, a relationship was found between the game experiences of students and their gameplay performance. These findings provide useful insights for researchers and educators wishing to incorporate DGs into classroom learning.

4.1 Introduction

Over the years, Higher Education (HE) institutions have been investing in technological tools to improve the learning experiences of students, make content delivery more effective, and provide opportunities for the development of digital skills. Simultaneously, there has been shifting attention to the development of realistic learning scenarios where technologies are used to support HE students in developing the skills required in the workplace. Several technological tools are being explored and used for delivering learning content to students. These include immersive learning technologies like virtual reality, augmented reality and DGs. DGs are increasingly being used for teaching and training in different educational settings (*e.g.* Chon *et al.*, 2019; Oren *et al.*, 2021; Suescún *et al.*, 2018; Suzuki *et al.*, 2021). There is increasing support for their use in engineering education given the potential of DGs to complement current pedagogical practices (Bodn*ar et al.*, 2016; Kittur and Islam, 2021; Udeozor *et al.*, 2022).

The use of DGs for instructional purposes is usually referred to as Digital Game-Based Learning (DGBL). Bahadoorsingh *et al.* (2016) defined DGBL as an instructional method that integrates educational content into DGs with the goal of engaging learners. DGs are widely accepted to be powerful tools for teaching and training for several reasons. Educational games that successfully pair instructional content with relevant game features are most likely to engage learners in gameplay, leading to the acquisition of knowledge and skills specified in statements of intended learning outcomes (ILOs) (Garris *et al.*, 2002). DGBL is considered relevant in HE because of its ability to foster contextual and authentic learning, experiential learning, collaborative learning, and problem-based learning, and also due to its ability to provide adaptive and appropriate feedback to learners (Oren *et al.*, 2021).

Since DGs are non-traditional teaching tools, several studies have aimed at understanding the views and experiences of students regarding the use of DGs for learning (Beavis *et al.*, 2015; Franco-Mariscal *et al.*, 2015). It is believed that the perceptions and acceptance of new technologies by intended users play a crucial role in ensuring successful outcomes (Herzog & Katzlinger, 2011; McMorran *et al.*, 2017). Beavis *et al.* (2015) stressed the importance of understanding the views of students and their previous experiences with games for learning to ensure the effectiveness of game-based learning pedagogy. In the previous chapter, engineering students raised concerns regarding game experience interference with academic performance when assessed with DGs. However, it is unclear if and how the game experiences of students and their perceptions towards DGBL reflect on their use of and performance in the game-based learning environment. From an educational point of view, a good performance in a DGBL environment can be seen as being able to demonstrate mastery of the competencies specified in the statements of ILOs by successfully completing the game tasks. Performance in the game can be measured through the physical variables that can be established by the software in line with the specifications defined by the ILOs of the specialist subject area. A good performance in a well-designed educational game would suggest a good understanding and application of the taught concept, hence, evidence of the acquisition of the competencies that were intended to be learnt.

Although studies on the effectiveness of DGBL have received much attention, there is still limited literature looking at the psychology of learners (Lu & Lien, 2020). A good understanding of how the psychological status of learners affects their performance in DGBL is essential to the design of high-quality interactions and effective instructions in DGBL environments (Lu & Lien, 2020). The current chapter aims at providing empirical evidence of the relationship between previous game experiences, perceptions of learning DGs, and the performance of engineering students in an educational game. This study does not intend to predict the future performance of students based on their experiences and perceptions but to explain the association between these variables.

4.1.1 Perceptions, experience and performance of students in DGBL environments

The growing interest in DGBL has led to several studies investigating the perceptions of students towards games for learning. The outcomes of these studies have generally been positive, with the majority of students agreeing that games enhance learning (Bolliger *et al.*, 2015; Sevim-Cirak and Yıldırım, 2020; Thanasi-Boçe, 2020; Udeozor *et al.*, 2021). Yue and Tze (2015) found that 58% of 51 computing students who took part in their study agreed or strongly agreed that DGs are viable teaching tools for improving learning experiences. 81% of the 222 university students who took part in an English language course agreed that games offer the opportunity to experiment with knowledge, with only 11 students reporting not seeing any pedagogical advantage of DG for English language learning (Bolliger *et al.*, 2015). The results of the study in Chapter 3 of this thesis showed that engineering students believe that DGs would be effective for engineering education and that students were willing to adopt DGs for learning. These outcomes are not specific to only students in HE. Other studies involving primary and high school students also found that overall, students show enthusiasm toward the use of learning games (Franco-Mariscal *et al.*, 2015; Beavis *et al.*, 2015; Bourgonjon *et al.*, 2010).

Whereas several studies have examined links between the perceptions of students and their performance in educational DGs, only a handful of studies were found to examine the influence of prior gameplay experiences of students on their performance in educational DGs (Sun, Chou and Yu, 2022; Yang and Quadir, 2018). In one study, the influence of prior gameplay experience on the problem-solving performance in a DG was examined using data from 267 elementary school children (Sun *et al.*, 2022). Sun and colleagues reported significant differences in the mean test scores of students based on the number of years of gameplay. It was found that students with more extensive gaming experience performed generally better than those with limited gaming experience on a game-based problem-solving test. This difference in performance was attributed to the greater familiarity that experienced players have with multiple game genres, making it easier for these sets of players to adapt quickly to game scenarios. While this finding seems to justify the concerns raised by engineering students in the previous chapters, a contradictory finding was reported in a different study. Yang and Quadir (2018) examined the effect of the prior gaming experiences of students on learning performance and anxiety. Using data from 55 elementary school children, they found significantly higher improvements in test scores of students with limited online gaming experience compared to those who claimed to have more online gaming experience ($\text{Mean}_{\text{High}} = 3.85$, s.d. = 11.83; $\text{Mean}_{\text{Low}} = 10.90$, s.d. = 12; $p<0.05$). No significant difference was found in the degree of anxiety experienced by low and high online gaming experienced students when learning with an educational DG ($\text{Mean}_{\text{High}} = 1.97$, s.d. = 0.88; $\text{Mean}_{\text{Low}} = 2.05$, s.d. = 0.74; $p>0.05$). These findings suggest that prior gaming experience negatively influenced performance, however, given that multiple-choice tests external to the game were used for the assessment, it can be argued that the findings add little to understanding the relationships between gaming experience and performance in an educational game. Moreover, Yang and Quadir (2018) failed to provide pre and post-game test scores that would allow for a more critical examination of the differences between the two groups. Nonetheless, it is worth reiterating that the participants of these studies were elementary (primary) school children, whose data might differ from those of HE students, especially engineering students.

Understanding the relationship between gaming experience, perceptions towards DGBL and the performance of engineering students in educational DGs is essential to achieving pedagogical effectiveness. Squire (2006) argues that performance elicits knowledge in video games as learners learn by doing within the constraints of the game environment. He emphasised the need for educators to study the "design experiences" of students that would influence knowledge acquisition and problem-solving. So far to the knowledge of the author, only a few studies have investigated the relationship between perception and performance in DGs (*e.g.* Ninaus *et al.*, 2017; Lu and Lien, 2020), but none have looked at the relationship between gaming experience, perception and the performance of students from a HE viewpoint. While the link between experience, positive perceptions and performance of students in game-based environments is yet to be fully established, a substantial body of research has shown that there is a relationship between positive affect and cognitive performance (Ashby *et al.*, 1999; Yang *et al.*, 2013; Liu & Wang, 2014).

4.2 Theoretical Background and Hypothesis Development

This research is grounded in the neuropsychological theory of positive affect. The neuropsychological theory of positive affect and its influence on cognition by Ashby *et al.* (1999) posits that positive affect, the positive feelings and emotions of a person, is linked to the release of dopamine which influences performance in a range of cognitive tasks. This is to say that the more positive experience or feeling one has (of something), the better their cognitive performance. When applied to DGBL, it is thus the expectation that positive gaming experiences and perceptions towards DGs would influence performance in the DG environment. Consistent with this theoretical view, a few studies have examined the relationship between perceptions and outcomes (Ninaus *et al.*, 2017; Kleinlogel *et al.*, 2020; Lu and Lien, 2020). In their study, Ninaus *et al.* (2017) evaluated the effect of the acceptance of DGs for learning on learning success. With data from 32 primary school participants, they found that learners with higher perceptions of ease of use and usefulness of games performed better in the post-game test. Another study by Lu and Lien (2020) investigated the relationship between the perceptions of playing and learning in a game-based environment, and the self-efficacy of 362 primary school learners in the same environment. They found a positive relationship between the perceptions and self-efficacy of students in the game-based learning environment. Similar findings were also reported by Kleinlogel *et al.* (2020) in their study investigating the role of meta-perception on performance in public speaking with 132 university participants. While these studies were consistent in their outcomes, one study did not find links between the perceptions of 287 medical students towards the use of videos for learning and their performance in tests (Mahmud *et al.*, 2011). This finding was attributed to the academic discipline of the student participants with reference to similar outcomes in previous studies on medical students.

As the use of DGBL in HE is relatively in its early stages, more research is required to understand how students use the technology, and how to support learning in the DG environments through appropriate design and deployment practices. The limited number of studies on perceptions and performance in the game-based learning context, and the inconclusive results on the relationship between perceptions, experience and performance in HE, warrant further research. Furthermore, most of these previous studies used external assessment methods, such as questionnaires and test scores (Kang *et al.*, 2017) for performance assessment, and non-established questionnaires for perceptions evaluation. This raises concerns about the validity of the outcomes. The current study, therefore, aims to bridge this knowledge gap and go a step further to evaluate the perceptions of HE students using an established technology acceptance model. It also uses embedded gameplay log data that are considered highly valid and reliable for game-based performance assessment (Shute & Ke, 2012). Spearman's correlation analysis is performed to evaluate the relationship, if any, between the game experiences of students, their perceptions of learning with DGs, and their performance in an educational game. The following questions will be answered by this research:

> RQ1: What are the gameplay experiences of chemical engineering students and how do they perceive the use of DG for engineering education?

The experience and perceptions of students towards learning games are explored, however, since the current cohort of students in HE institutions is regarded as digital natives (Prensky, 2001), it is expected that the majority of the participants will report positive experiences and perceptions.

> RQ2: Are there any relationships between game experiences, perceptions, and gameplay performance of engineering students?

Consistent with the neuropsychological theory of the influence of positive affect on performance, it is expected that:

> H1*: Positive experiences with DGs will positively correlate with the perceptions of students towards learning DGs and their performance in the educational DG.*

4.3 Research Methodology

In order to establish whether any correlations exist between game experience, perceptions and performance in a DG, a correlational research design was employed (Cohen *et al.*, 2007). Correlational research design is often used to predict or explain relationships among variables. Two types of correlational research designs commonly used for these are the explanatory design and the prediction design (Creswell, 2011). Explanatory (sometimes referred to as "relational") design is used when the objective of the study is to explain the association between or among variables. Since the aim of this chapter is to understand associations between three variables, the explanatory design was used. However, when the goal of a study is to identify variables that will predict an outcome, the prediction correlational research design is recommended (Creswell, 2011).

To determine the performance of students in the game using log data, cluster analysis was carried out. Cluster analysis is an unsupervised machine learning classification technique used to group similar and homogenous sub-samples of cases or people (Cohen *et al.*, 2017). Useful for discovering groups in data, it measures distance, proximity and similarity between samples. A 2-step cluster analysis involving hierarchical and K-Means clustering were used to partition students into groups based on their performance on game tasks. A hierarchical clustering technique is often used to explore the number of classes or clusters in data. It utilises divisive or agglomerative methods to separate individuals into similar groups represented in a dendrogram (Landau *et al.*, 2011). Agglomerative or bottom-up clustering finds patterns in data by first assigning each data point to individual clusters before merging close clusters by calculating the distance between them. In the end, all data points are merged into one cluster. Divisive clustering goes the reverse way, it first merges all data into a single cluster before splitting them into individual clusters. These processes of partitioning data are represented in the dendrogram which is a tree-like diagram of the data points showing nodes where clusters merge.

Agglomerative clustering methods are the most commonly used hierarchical clustering technique in technical and non-technical publications. Although as effective as divisive clustering, agglomerative clustering is preferred over divisive clustering because of the high computational requirements needed for the latter (Kumar *et al.*, 2019). For this reason, an agglomerative clustering method was used in the current study. Some of the linkage methods used for agglomerative clustering include single linkage, complete linkage, average linkage and Ward's method (Landau *et al.*, 2011). These are used for grouping similar clusters. There are no specific conditions for choosing a given method over the other, as each method achieves similar results using different approaches. Ward's method was chosen for the current study. Ward's method partitions data by computing variance rather than the distance between clusters unlike the other methods (Kumar *et al.*, 2019). It is known to be conservative and performs significantly better than the other clustering methods (Eszergár-Kiss & Caesar, 2017; Landau *et al.*, 2011). While hierarchical clustering was used to identify the number of clusters in data, K-Means clustering was used to partition people into a fixed *k* number of clusters based on similarities and often measured by the distance between the data and the cluster mean (Slimani *et al.*, 2018). This method of classification is widely used, for example, for weather classification, astronomy and market research. In recent studies, it was also explored as a means of assessing the performance of students from game log data (Kerr and Chung, 2012; Lin, Hsieh, Hou and Wang, 2019). K-Means clustering is used in the current study to partition data on the performance of students into the pre-determined number of groups identified in the hierarchical clustering. IBM SPSS version 27 was used for these analyses.

Whereas cluster analyses were used to identify performance clusters in the data, correlational analysis was used to identify relationships between the three variables of interest in this chapter. Correlation analysis is a statistical analysis used to measure the association or relationship between two or more quantitative variables. A correlation coefficient which is a single value between -1 and 1 is produced at the end of the analysis to establish the strength of the association between the variables (Gogtay & Thatte, 2017). There are parametric and non-parametric measures of correlations such as Pearson's correlation and Spearman's correlation, respectively. Depending on the type and normality of data used for the analysis, either method can be used. For Pearson's correlation coefficient, variables must be interval or ratio, and the data must be normally distributed. Spearman's correlation was therefore chosen over Pearson's correlation for this study because some of the data in this study failed to meet all the assumptions of Pearson's correlation (Gogtay & Thatte, 2017; Khamis, 2008). The ordinal nature of the game enjoyment and perception variables indicated that the non-parametric Spearman's rank correlation was most appropriate. This analysis was automatically computed on IBM SPSS version 27.

4.3.1 Research participants

The data of a subset of the 125 students who participated in the previous study detailed in Chapter 3 were used for this study. Only the questionnaire responses of 58 students that could be matched to their gameplay data were used for this study. During the data collection periods, students were asked to complete the questionnaire before playing the game. As data were collected remotely during the COVID-19 pandemic, some students skipped the questionnaire and only played the game, while in other cases, technical challenges made it impossible to locate the corresponding gameplay data and questionnaire responses of students. These resulted in the use of data from a total of 58 chemical engineering students from KU Leuven and Imperial College, London. There were 39 male and 18 female students in this study and 97% of them were between the ages of 20 and 29 years.

Table 4.1: Demographic information of (n=58) engineering students who took part in the correlational study.

		Absolute Frequency
Gender	Male	39
	Female	18
	Unspecified	1
Age	Under 20 years old	2
	20-29 years old	56
Prior gameplay	Yes	54
	No	4
Gaming habits (per week)	Less than 5 hours	25
	5-10 hours	21
	11-20 hours	7
	21-30 hours	4
	31 hours or more	1

4.3.2 Procedure

The study took place remotely during the COVID-19 restrictions. The study required participants to complete an online questionnaire before playing an educational DG known as CosmiClean. Participants were sent PowerPoint presentations of the study goals, gameplay instructions and a link to the game. Once the game was downloaded by clicking the link, students were prompted to complete the online questionnaire before returning to the gameplay. Participants were given three weeks to play the game and were requested to complete a minimum of 25 levels. This extended study time was given to provide enough time for students to take part in the study during term time. Additionally, it would potentially accommodate students with different gaming skills and learning preferences, giving them enough time to navigate the game and complete the required minimum number of levels.

4.3.3 Data collection

The online questionnaire adapted from the UTAUT 2 (Venkatesh *et al*., 2012), detailed in Chapter 3 of this thesis, was used for data collection. Ethical approval was granted for this study. Participants were informed about the purpose of the study and how their data will be used. They were asked to use the most optimal solutions to complete the game tasks and were also informed that the log data of their performance would be analysed. All data collected were anonymised and cannot be traced back to the participants. Participants responded to a total of 20 6-point Likert scale questions that measured the perceptions of the use of DGs for engineering education on six constructs shown in table 3.2. Demographic information and details of the prior gaming experiences of participants were also collected. For gameplay performance, log files of the CosmiClean gameplay of participants were retrieved.

4.3.4 CosmiClean game

CosmiClean (https://recyclegame.eu/the-game/) was launched in 2018 as an educational game with strategy, puzzle, and adventure elements. It is a product of the European Institute of Innovation & Technology – Knowledge and Innovation Community (EIT-KIC Raw Materials) game project of the European Union Horizon 2020 (EUH2020) and was designed by LuGus Studio, Belgium. The game was designed to expose players to the science and challenges of waste separation and recycling using principles of automated industrial processes. With 56 levels of increasingly challenging gameplay, CosmiClean uses high-quality graphics and an engaging narrative to provide a fun game-based learning experience.

The game places the player in the position of an artificial intelligence (AI) trying to save a stranded damaged spaceship carrying a ton of waste across the galaxy. With no resources available to repair the ship, the AI resorts to using the waste cargo. The player has to sort the waste materials and use them to repair parts of the damaged ship as illustrated in figure 4.1. With a working 3D printer on the ship, the player is also able to design and print different separators needed to sort the waste.

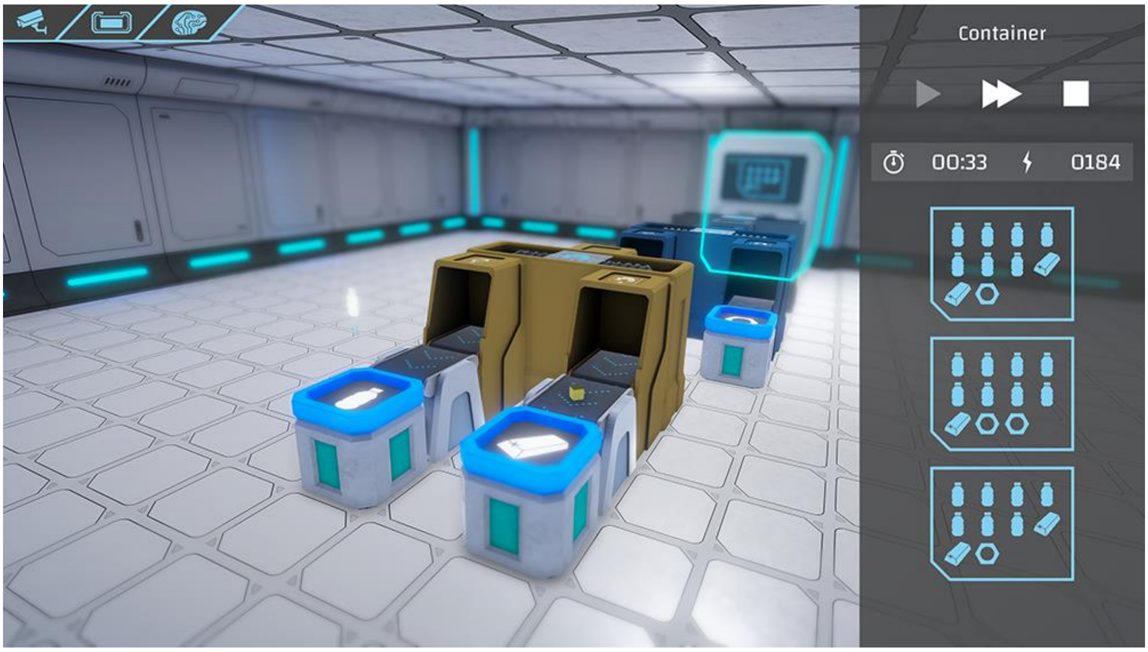


Figure 4.1: Screenshot of a separation process in CosmiClean

When viewed from an engineering education point, CosmiClean teaches the heuristics of separation processes through waste recycling. The players have to determine and use appropriate resources (including energy and processors) in the correct sequence to separate mixtures of waste by exploiting different properties of the materials at each level of the game. Although designed for the wider public, the gameplay tasks are closely aligned with some core modules of chemical engineering education, such as the principles of separation operations and the elementary principles of project-based plant design. In the game, players make strategic decisions on which processors (i.e. equipment) to use, optimal sequencing, as well as the configuration of each processor to ensure the efficiency of the separation and the recovery of the raw materials from the waste cargo. This game was selected for this study because it was designed by engineering experts with learning outcomes related to chemical engineering. Its design was also recognised as outstanding and was given the Comenius-EduMedia Siegel award in 2019 (https://comenius-award.de/) so the quality of the learning experience with the game is already recognised in the community.

4.3.5 Data analysis

Data analyses were carried out using IBM SPSS version 27 and RStudio 4.0. To answer RQ 1, a descriptive analysis of the questionnaire responses of students was conducted using SPSS. For RQ 2, log files from the gameplay of students were processed in RStudio before being analysed in SPSS. The gameplay performance of students was evaluated using a 2-step clustering technique: a hierarchical clustering followed by a K-Means clustering. The log data used for this analysis are the average time on task (ATT), the number of levels solved (LS), and the total energy expended (TEE), that is, energy used to complete the levels of the game. These metrics are captured during gameplay. A sample of the log data of the gameplay activities of students is presented in Appendix 4. A.

To assess performance in the game, the gameplay log files of participants were retrieved and processed in Rstudio 4.0 to extract relevant data while the analysis of the data was carried out in IBM SPSS version 27. A descriptive analysis of the gameplay actions of the students is presented in table 4.2. Data were screened so that only the data of students who completed the questionnaire and played at least ten levels of the game during the given time frame were included in this analysis. The data of 14 students who solved less than ten levels were excluded because they would add little to no information on the gameplay performance characteristics of the students and could potentially impact the outcome of the correlation analysis.

Table 4.2: Descriptive data of the performance of (n=44) engineering students in CosmiClean.

	Minimum	**Maximum**	**Mean**	**Std. Deviation**
Levels Solved **(LS)**	10	57	29	14
Average Time on Task (mm:ss) **(ATT)**	01:00	11:29	02:19	01:46
Total Energy Expended **(TEE)**	359	11224	1815	1885

Hierarchical clustering of the performance of students based on LS, ATT, and TEE was first carried out. Before this analysis, checks were carried out to confirm that the recommended minimum sample size requirement was met. Sample size justifications are rarely given in studies performing cluster analysis (May & Looney, 2020) and there are several opinions on what the minimum sample size for cluster analysis should be, with no established criteria. Also, discussions on this topic have been limited to specific fields such as in business for market segmentation (Dolnicar *et al.*, 2016) and biomedical research (Dalmaijer *et al.*, 2022). Formann’s sample size recommendation based on goodness-of-fit testing using Chi-square tests is generally applicable to most domains for sample size estimation (Formann (1984), cited in Dolnicar *et al.*, 2016). This sample size estimation was adopted for this study. According to Formann, a minimum sample size of five times the value of two to the power of the number of variables in the segmentation base, shown in equation

4.1, is recommended. With three variables for the current analysis, the recommended minimum sample size of 40 is met, given that the sample size of this study is 44.

$$n_{\min} = 5 \times 2^{\chi} \qquad \text{Equation 4.1}$$

$$= 5 \times 2^{3} = 40$$

where $n_{\min}$ is the minimum sample size; χ is the number of variables.

To begin the analysis, the scores on the variables were first standardised to make the relative weight of each variable equal. From the dendrogram in figure 4.2, the longest branch at 25 offers the best separation between the merged clusters. It is recommended to “cut” dendrograms at the similarity point corresponding to the longest branch (length of the horizontal lines) to obtain unique clusters (Forina *et al.*, 2002). The decision on the optimal number of clusters from a dendrogram diagram is often subjectively made (Unglert *et al.*, 2016). However, the choice of two rather than three clusters from figure 4.2 was made because the third potential cluster offers no additional information as it contains only one case/student. Therefore from figure 4.2, the optimal number of clusters in the data is said to be two.

4.4 Results

4.4.1 Game experiences and perceptions of engineering students towards DGBL

The data collected from the questionnaire indicate that 93% of the participants have played some sort of game before. As shown in figure 4.3, 43% of the students reported spending less than 5 hours a week while 2% spend over 31 hours a week on gameplay. Participants were asked to rate their enjoyment of digital gameplay on a scale of 1 to 10. As detailed in figure 4.4, 69% of the participants rated their enjoyment of games highly, scoring those between 7 and 10.

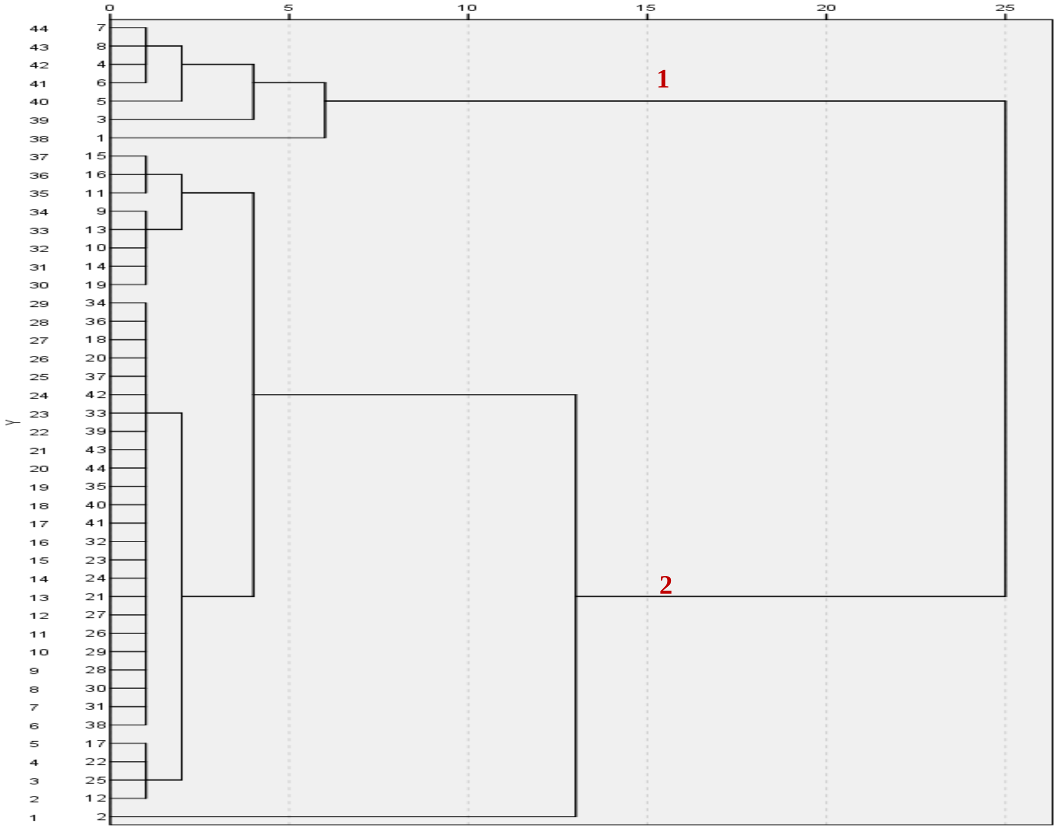


Figure 4.2**:** Dendrogram of log data using Ward’s Method, with (n=44) engineering students who took part in the correlation study.

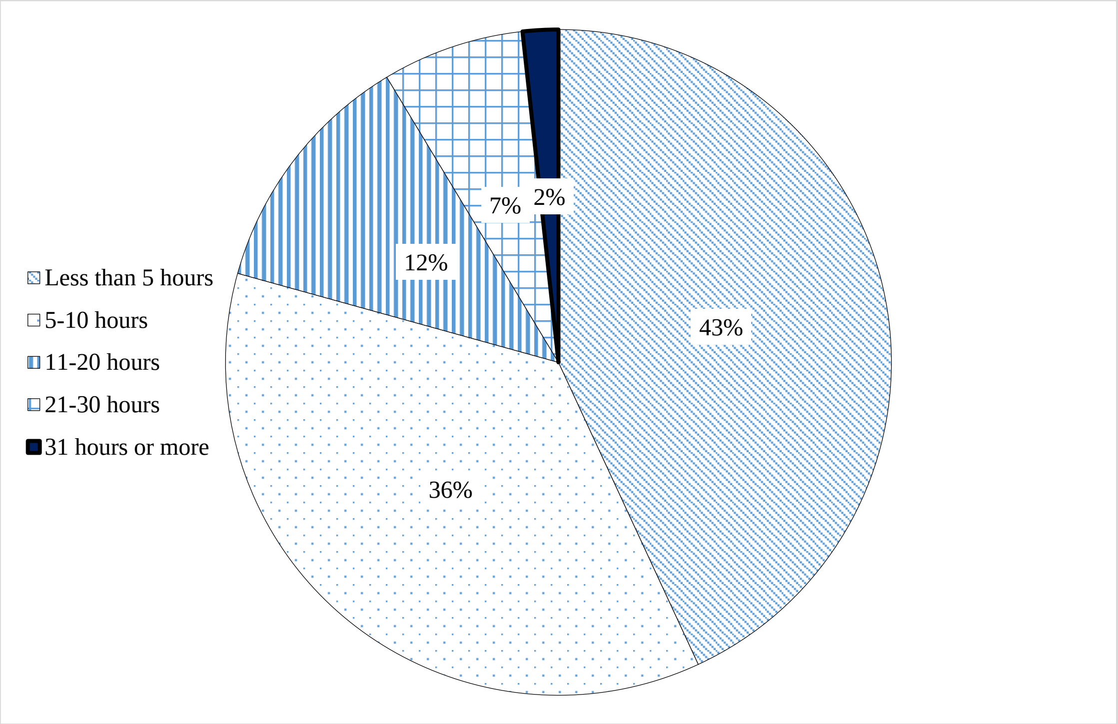


Figure 4.3: Weekly gameplay duration of (n=58) engineering students.

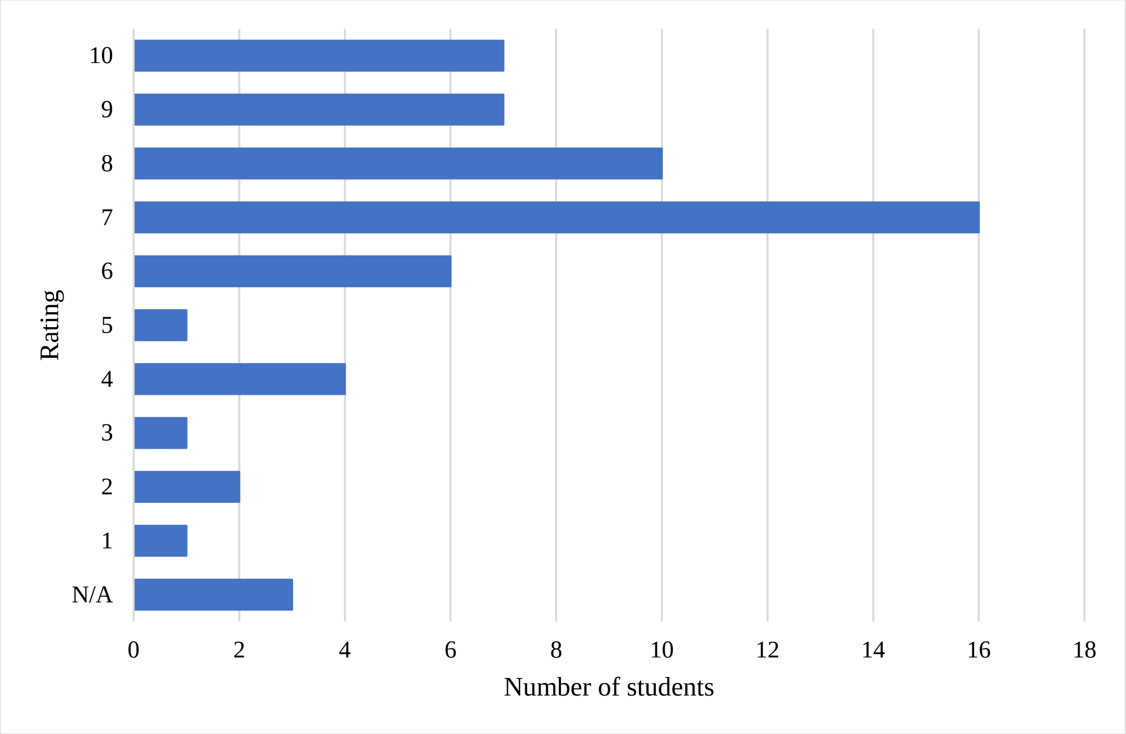


Figure 4.4: Game enjoyment ratings of (n=58) engineering students who took part in the correlational study.

With a subset of the survey data of students analysed in Chapter 3 of this thesis used for this analysis, and the reliability and validity checks of the questionnaire items confirmed in Chapter 3, only a descriptive analysis of the responses is detailed here. The mean scores of participants on all measured constructs were calculated and are presented in table 4.3. The overall mean ratings on all constructs were relatively high, ranging between 3.8 to 4.8 out of 6. Given that 93% of the participants reported previous gameplays, with high ratings on the game enjoyment measure, and relatively high perceptions of DGBL, it can be inferred that chemical engineering students have highly positive experiences and perceptions of learning with DGs.

Table 4.3: Mean rating of n = 58 engineering students on their perceptions of the use of DGs for education measured on six UTAUT 2 constructs.

Constructs	**Mean**	**Std. Deviation**
Performance Expectancy	4.53	0.70
Effort Expectancy	4.77	0.80
Facilitating Condition	3.78	1.09
Social Influence	4.76	0.70
Hedonic Motivation	4.10	0.92
Behavioural Intentions	3.85	0.87

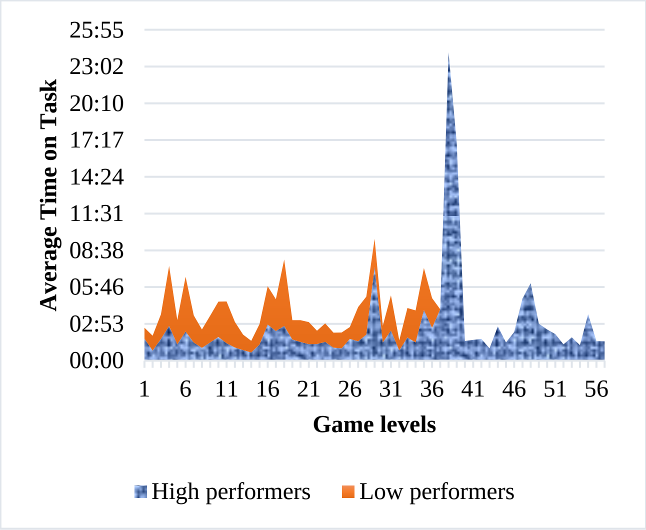
4.4.2 *Relationship between experience, perceptions and gameplay performance*

To evaluate the relationship between game experience, perceptions toward education DGs and performance in CosmiClean, the performance of students in the game was first determined. In subsection 4.3.5, two potential clusters of performance were in the data based on the outcome of the hierarchical clustering carried out. Following that outcome, a 2-cluster K-Means analysis was performed to identify the 2 groups of students from the gameplay data. From the results presented in table 4.4, the students in Cluster 1 performed better than those in Cluster 2, and are labelled the “High” performers and “Low” performers, respectively. Although both groups look comparable on the variable average time on task (ATT), Cluster 1 students can be said to have performed better. These sets of students played more levels of the game. It is worth mentioning that the performance of students on the game, measured by TEE, ATT and LS could have been influenced by their learning styles, English language proficiencies and neurodiversity. However, these metrics were the only relevant metrics captured by the game and thus they were used here for research purposes.

Table 4.4: Characteristics of the identified clusters of students following a 2-cluster K-Means cluster analysis of the data of (n = 44) engineering students.

	Cluster 1-Higher performers		**Cluster 2-Lower performer**	
	Mean	S.D	Mean	S.D
Levels Solved **(LS)**	54.25	7.778	23.14	6.617
Average Time on Task (mm:ss) **(ATT)**	02:33	01:56	02:16	01:46
Total Energy Expended **(TEE)**	4764.38	2862.504	1159.17	543.343
Number of students	**N=8**		**N=36**	

The game is designed to be incrementally challenging and complex as player progress through the levels. This means that for higher levels of the game, it is expected that students spend more time completing the tasks. It is also expected that the average time spent on each level of gameplay by a student who solved 50 levels of the game would be longer than the average time spent by a student who solved 20 levels of the game. As shown in figures 4.5 and 4.6, Cluster 1 students outperformed students in Cluster 2, given that on each level of the game, students in Cluster 1 (*i.e.* the “High” performers) completed more levels of the game, spent comparatively less time on each level of the game, and used less energy. The comparative result on the mean scores on ATT shown in table 4.4 is due to the fewer levels solved by students in Cluster 2 which affected the overall gameplay time and the computed average time spent on each level.



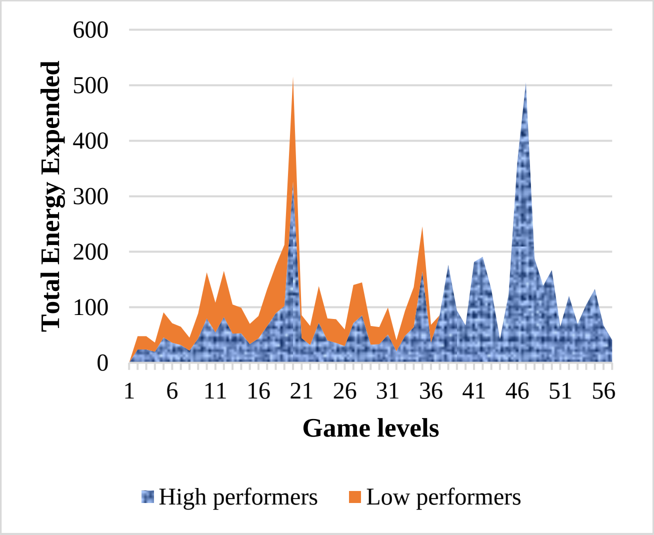


Figure 4.5: Average time per level solved.

Figure 4.6: Total energy expended per level.

With the performance of students in CosmiClean determined, Spearman's correlation analysis was carried out to explore the relationships between game experience, perceptions and performance. The mean scores of students on all measured constructs of the modified UTAUT 2 model were used for the perception variable. Game experiences were measured on three variables – "prior gameplay", *i.e.* whether they had played games before, "gaming habits", the average time students spend on gameplay per week, and their "game enjoyment" ratings. The minimum sample size for this analysis based on Fisher z-Transformation was found to be 30, given a statistical power of 0.8, an alpha level of 0.05 and a true correlation of 0.5 (May & Looney, 2020). A sample size of 44 used for this study implies that the minimum sample size requirement for this analysis was met.

The relationship between the gameplay performance clusters, perceptions of students, and their experiences with DGs was first explored. From table 4.5, no significant relationship was found between the clusters of students, and their perceptions and game experiences. This is evident by the 2-tailed significance values that are greater than 0.05 ($p>0.05$). Spearman's correlation coefficient, Spearman's rho, has a value between -1 and 1, indicating the strength of the relationship between the two variables. The higher the value, the stronger the correlation between the variables. A negative value indicates a negative correlation while a positive value suggests a positive correlation between two variables. To ascertain whether a negative or positive correlation is significant, *i.e,* unlikely due to chance, the statistical significance of the result (*p*-value) is checked. Although no significant relationship was found between cluster performance, perceptions and experiences of students, the negative correlations shown in table 4.5 imply that the higher the cluster value (lower performers) the lower they rated the questions on game experiences and perceptions.

Table 4.5: Spearman’s correlation coefficients for relationships between gameplay performance clusters, game experience and perceptions of (n = 44) engineering students. Where rho is the correlation coefficient represented by values of -1 to 1 and sig. is the statistical significance values (p-values) represented by values of 0 to 1.

		Performance clusters	**Prior gameplay**	**Game Enjoyment**	**Gaming habits**	**Perceptions**
Performance clusters	Spearman’s rho	1	-0.180	-0.066	-0.231	-0.102
	Sig. (2-tailed)(p)		0.242	0.669	0.131	0.510
	N	44				

Next, further analysis was carried out to determine whether any relationship exists between the game experiences, the perceptions of students and their overall gameplay performance measured by the three variables considered. As depicted in table 4.6, the correlation analyses showed moderate to high significant relationships (Cohen, 1992) between the gaming habits of students and their game enjoyment ratings ($r = 0.648$), between LS and TEE ($r = 0.868$), which are significant at $p < 0.01$, and between game enjoyment and ATT ($r = -0.318$), significant at $p<0.05$. These suggest that students who generally spent longer hours per week on gameplay rated their gameplay enjoyment higher. The negative relationship between game enjoyment and ATT indicates that students who rated their enjoyment of games highly were faster at completing the game tasks, that is, they spent the least amount of time solving each level of the game. Nonetheless, no significant relationship was found between gaming habits and ATT ($r = -0.176$, $p = 0.253$), nor between any variable on gameplay performance and prior gameplay. No relationship was found between the perceptions of students and their gameplay performance. Therefore, based on these findings, hypothesis 2 is partially accepted.

4.5 Discussions and Implications

The goal of this chapter is to first explore the perceptions and experiences of engineering students toward DGBL and then investigate the relationships between these and their gameplay performance in an educational DG. To accomplish these goals, data from 58 chemical engineering students who participated in the study presented in Chapter 3 were used. The majority of the students reported having prior entertainment digital gameplay experience, spending varying lengths of time per week on gameplay. It was apparent from the responses that most of the undergraduate engineering students that were surveyed enjoy playing DGs and have positive perceptions of the use of DGs for engineering education. 69% of the students rated their enjoyment of DGs high with ratings of 7 to 10 out of 10, and 81% of the students scored an average of 4 to 6 out of 6 on the statements measuring perceptions. These resonate with the studies of Bolliger *et al.* (2015) and Cook-Chennault and Villanueva (2020) that found that over 76% of HE students and a majority of engineering students enjoyed playing games. Although the participants of the current study claimed to enjoy playing games, this did not necessarily reflect on their gameplay performance.

Two clusters of performance were identified in the gameplay data, but only 18% of the students were categorised as High performers while the rest of the students were labelled as Low performers. The High performers completed more levels of the game, spent less time solving each game level, and used less energy per level solved, compared to the Low performers. Backed by the proven relationship between positive affect and cognitive outcomes (Ashby *et al.*, 1999), and given the overwhelmingly positive views and prior game experiences reported by the majority of the students, it was expected that the performance of a similar proportion of the students would be considered as High. This was not the case as non-significant correlations were found between the performance of the two groups of students and their game experiences and their perceptions of learning games. This result was found to contradict previous related studies (Ninaus *et al.*, 2017; Kleinlogel *et al.*, 2020; Lu and Lien, 2020) that are discussed in detail below. However, the findings of the current study might have been affected by the variables used for the analysis, hence further examination of the relationships between the three variables was carried out, this time without the clusters.

Although no relationships were found between performance, game experiences and perceptions at the performance cluster level, some positive relationships were identified when performance was measured by LS, ATT and TEE rather than Cluster 1 (High performers) and Cluster 2 (Low performers). Within the limitations of the statistical significance of the data collected, the results indicated a medium significant relationship between game enjoyment scores and the average time on task (ATT), the only significant relationship found between the game experience and performance. This finding supports the neuropsychological theory of positive affect and cognition (Ashby *et al.*, 1999), considering the moderate negative correlation between game experiences (game enjoyment) and performance (time on task). Put differently, a moderate positive correlation between game experiences and task completion times was found. This finding implies that technology-savvy students and those who enjoy gameplay will likely perform better in educational DGs. Educators would need to be able to account for this effect when implementing educational DGs for teaching, learning, and especially for assessment.

The unexpected result showing no correlation between the performance of students in the game and their perceptions of DGBL contradicts prior findings (*e.g.* Lu and Lien, 2020; Ninaus *et al.*, 2017). These differences could be due to the population sampled in the studies. The studies of both Lu & Lien and Ninaus *et al* measured perceptions of primary-aged pupils, whereas, in this study, adult learners in HE were sampled. This could also be the reason why the current finding aligns with that of Mahmud *et al.* (2011) who found no relationship between perceptions of the use of videos for learning and the test performance of medical students. While the ages of the participants of the current study may have played a role in the outcomes observed, the measurement instrument used to measure perceptions would have also played a part. The responses to questionnaires are often prone to biases inherent in self-reports (Herde *et al.*, 2019). Measuring the opinions of students based on their responses to a questionnaire that was not specifically designed to measure perceptions towards educational DGs is likely to influence the results obtained. In previous studies, modified versions of existing technology acceptance models were used to measure perceptions toward educational DGs which could have impacted the reported results. As emphasised in Chapter 3, there is therefore the need for research into theoretical models and questionnaires for assessing the perceptions of students towards immersive learning technologies. Nevertheless, this finding weakens the argument that the perceptions of students toward game-based learning have a significant impact on performance and effectiveness. The implication of this is that perceptions, attitudes, or expressed interest in games may not be sufficient to ensure that students will use educational DGs as expected.

Table 4.6: Spearman’s correlation coefficients for relationships between gameplay performance, game experience and perceptions (n = 44) engineering students. Where rho is the correlation coefficient and sig. is the statistical significance values (*p*-values) represented by values of 0 to 1 with *p*<0.05 indicating statistical significance of the highlighted correlation coefficients.

		Prior gameplay	**Game enjoyment**	**Gaming habits**	**Perceptions**	**Levels Solved**	**Average Time on Task**	**Total Energy Expended**
Prior gameplay	Spearman’s rho	1	-0.079	-0.096	-0.138	0.190	0.215	0.180
	Sig. (2-tailed)		0.611	0.537	0.373	0.216	0.161	0.241
Game Enjoyment	Spearman’s rho		1	**0.648****	.017	0.126	**-0.318***	0.193
	Sig. (2-tailed)			**0.000**	0.913	0.417	**0.035**	0.209
Gaming Habits	Spearman’s rho		0.648**	1	-0.017	0.191	-0.176	0.206
	Sig. (2-tailed)		0.000		0.914	0.215	0.253	0.180
Perceptions	Spearman’s rho		0.017	-0.017	1	-0.090	0.081	-0.091
	Sig. (2-tailed)		0.913	0.914		0.563	0.601	0.558
Levels Solved (LS)	Spearman’s rho		0.126	0.191	-0.090	1	-0.063	0.868**
	Sig. (2-tailed)		0.417	0.215	0.563		0.683	0.000
Average Time on Task (ATT)	Spearman’s rho		-0.318*	-0.176	0.081	0.081	1	-0.183
	Sig. (2-tailed)		0.035	0.253	0.601	0.601		0.235
Total Energy Expended (TEE)	Spearman’s rho		0.193	0.206	-0.091	**0.868****	-0.183	1
	Sig. (2-tailed)		0.209	0.180	0.558	**0.000**	0.235	
	N		44	44	44	44	44	44
	Correlation is significant at the **p< 0.01 **; p<0.05 *** (2-tailed) shown in BOLD							

This might, in turn, have negative effects on the expected learning gains. Certainly not to trivialise the importance of understanding the views of students towards DGBL, this study emphasises the need for research into factors that could positively influence cognitive outcomes in DGBL and not simply rely on the initial perceptions of students when designing or using DGs for teaching. The significant relationships found between some game experience variables and gameplay performance variables indicate that it is worth paying more attention to the objectively measurable experiences of students as opposed to self-reported perceptions which may not reflect the actual views of students for several reasons.

While statistically non-significant, the positive correlation between prior gaming and gameplay performance suggests that the performance of students who had played DGs previously did not significantly vary from those who had not played DGs before. Aligned with the findings of Yang and Quadir (2018) but contradicting Sun *et al.* (2022) who similarly used educational games in their studies, this outcome could be because of the statistical limitations posed by the data of the students studied. Yang and Quadir found that the students with low gaming experiences paid better attention to the task description and instructions, whereas, the students with higher gaming experience approached the gameplay like they usually do, paying little attention to the task descriptions given. Given that almost all the students (95%) in the current study indicated that they had previously played DGs, the results of the statistical analysis are bound to be skewed. Additionally, as CosmiClean is a well-designed educational game that applies appropriate learning principles such as scaffolding measures and hints (Shute, Ke, and Wang, 2017), the performance of students is expected to be relatively similar regardless of prior gameplay experience. Nevertheless, the data is representative of the cohorts of students in HE institutions today. Limited by the sample size of this study, these findings provide some insights into the possible relationship between prior gameplay experiences of students, their perceptions of educational DGs and their performance in an education DG. For educators interested in DGBL, it would be worth understanding the previous gameplay experiences of students to look out for possible interference with their use of and the effectiveness of educational DGs.

4.6 Limitations

One limitation of this research is the sampling method and the small sample size of the participants which limit the generalisability of the findings. Future research should consider repeating this study using a larger sample size. The non-significant correlation between gameplay experience and performance could have been further investigated using the Tukey-

Kramer method (Games *et al.*, 1981) provided the variances in the variables are homogeneous. Future studies should consider repeating this study with a larger sample size and an equal number of gamers and non-gamers, or consider appropriate comparison tests for non-equal sample sizes. Furthermore, the elimination of the data of the 14 participants who completed less than 10 levels of the game could have impacted the outcome of the analysis of this chapter. Future studies should consider more appropriate methods of handling missing data such as pairwise deletion or mean substitution to minimise bias (Kang, 2013).

The questionnaire used to measure the perceptions of students towards DGBL in the current study was adapted from an existing technology acceptance model and may not have captured all relevant factors affecting the use of DGs for learning as found in Chapter 3 of this thesis. Learning styles, gameplay experiences and relevance to the module or major of the students are some constructs to consider incorporating into a technology acceptance model relevant to immersive learning technologies. A few studies have considered altering existing technology acceptance models for game-based learning (Bourgonjon *et al.*, 2010; Hsu and Lu, 2004; Shiue, Hsu and Liang, 2017). These authors introduced attitude, game experience, learning opportunities, and flow experience to existing models. Bourgonjon *et al.* (2010) evaluated perceptions of video games for classroom use using the modified model but the results showed lower predictive power of the model compared to the UTAUT 2 model. Shiue and colleagues, on the other hand, failed to indicate the predictive power of their model while the model of Hsu and Lu (2004) is intended to measure perceptions to play online games making them unsuitable for the current chapter.

4.7 Conclusions and Future Studies

This chapter highlights the need for more research into the influence of the experiences and perceptions of students on their performance in educational DGs. Despite the significance of some of the findings of this chapter, it is worth emphasising that the results must be interpreted cautiously given the limitations posed by the small sample size of the participants of this study. Future studies should however consider repeating this study with larger sample sizes. So far, no known validated questionnaire with good predictive power designed specifically to evaluate the perceptions or acceptance of game-based learning from the viewpoint of students exists. Future studies should develop models relevant to immersive learning, building on the views and experiences of students, previous findings, and relevant theoretical frameworks. Finally, the results obtained in this current study begin to raise questions on how to effectively measure learning in immersive environments. As the experiences of students with different immersive technologies, especially VR and AR, would likely differ, research is required to understand how students interact in immersive learning environments and how this could potentially affect the assessment of performance in these environments. Furthermore, there is a need to explore how different learners, particularly, neurodivergent learners interact with different educational game genres in order to design immersive learning approaches that take into consideration the needs of these learners.

4.8 Note:

Part of this research is published in the **Journal of Educational Computing Research.** Udeozor, C., Russo Abegão, and J. Glassey (2021). An Evaluation of the Relationship Between Perceptions and Performance of Students in a Serious Game. *Journal of Educational Computing Research*, *60*(2), 322-351 doi:10.1177/ 07356331211036989.

Chapter 5. Gameplay Behavioural Patterns and Performance of Engineering Students in an Educational Digital Game

In the previous chapters, the applications of immersive learning technologies in engineering education and the views of students towards the use of digital games (DGs) for learning and assessments were explored. The resistance and concerns of engineering students regarding the use of DGs for assessments were highlighted by the results of the study in Chapter 4. The findings of that chapter also raise questions about measuring learning performance when using immersive technologies for pedagogy. An understanding of how students behave within immersive learning environments is likely to provide valuable information that could inform assessment designs as well as other pedagogical practices. Although a few studies have attempted to provide insights into the behavioural patterns of students in game environments, the studies either sampled younger learners or used entertainment DGs (Hou, 2012, 2015). There remains a gap in the literature on the behavioural patterns of higher education (HE) students in immersive learning environments. This chapter uses grounded theory research design to provide an understanding of how students behave in educational DGs. Following an inductive approach, log data from the gameplay of engineering students are analysed for behavioural patterns and an explanation is developed for the patterns observed.

Furthermore, this study explores the use of log data for game-based assessment. As observed in Chapter 1, assessments with immersive learning technologies are predominantly based on external assessment methods, regardless of the potential of these technologies to enhance authentic performance assessments. While a few studies have utilised embedded assessments to measure the performance of students on general competencies, limited studies reported the use of game log data to measure domain-specific competencies. This chapter, therefore, presents an exploration of game log data for understanding the behaviours and performance of 58 engineering students in an educational DG. Sequential behavioural pattern analysis and cluster analyses were used for the analyses. The findings of this study highlight important characteristics of the behaviours and gameplay strategies of engineering students in DG environments. These findings would be particularly useful to educators and researchers in the field of immersive learning.

5.1 Introduction

The field of human-computer interaction (HCI) has grown rapidly as computer technology becomes a significant part of everyday life. The need to understand human behaviours in these environments in order to inform theory creation, the design of computers and computer programs, and the applications of these technologies to businesses, education and government, is of particular interest to computer scientists (Norman, 2017). Similarly, the entertainment industry invests hugely in research into player behaviours to help improve the gameplay experiences of players (Bakkes *et al.*, 2012). In education, substantial efforts are being made to understand human cognitive behaviours with theories such as behaviourism, constructivism, experiential learning theories and others, emerging from research to inform effective learning and teaching (L&T) practices. As computers become an integral part of L&T in schools, conversations on digital learning theories are emerging (Martin & Betrus, 2019). The attention that digital game-based learning (DGBL) has received over the years has also led to research into learning psychology, theories and their impacts on DGBL (Boyle, Connolly and Hainey, 2011*;* Wu *et al.*, 2012). Some researchers explored the psychology of DGBL through the lenses of established learning theories (Wu *et al.*, 2012), while others have proposed new principles for designing and evaluating DGs (Kiili, 2005). Regardless of the psychological angle through which DGBL is viewed, it is believed that learning processes are influenced by several factors including, but not limited to, the learning environment (Austin *et al.*, 2001).

5.1.1 Gameplay behaviours

Understanding the behaviours of students in technology-aided learning environments is crucial for informing effective pedagogical practices. Behavioural analysis of the activities of students in DG environments provides an in-depth understanding of their learning process (Hou, 2015), information that could enhance the design of relevant pedagogical methodologies. DG learning environments differ from traditional classrooms, workshops and laboratories. Game narratives, mechanics and other defining elements of games influence behaviours in games. Mapping the behaviours of players in entertainment games is commonplace as game developers seek to improve player experiences and satisfaction using artificial intelligence (AI) (Bakkes *et al.*, 2012). In education, however, extensive research on the effectiveness of DGs has been conducted, but limited research exists on the behaviours of students in DGs and how these affect pedagogy. A few studies such as that by Westera *et al.* (2014) used the game log files of students to identify gaming behaviours of students based on the time spent on different game activities. Regression analysis was used to determine the relationship between the gaming behaviours and post-test scores of students. The study found that video access rates and overall activity rates of students in the game were good predictors of learning efficiency. In other studies, the actions of students were sequentially mapped in order to identify the different types of learners, as well as gender differences inherent in the gameplay data (Hou, 2012, 2015). While one study assessed behaviours in an online game (Hou, 2012), the other utilised a role-playing simulation game to distinguish different types of players (Hou, 2015). The contributions of these studies are valuable to the field of DGBL, however, they are limited in scope and approach, and consequently in their wider application in the field.

Behavioural analysis of the educational gameplay interactions of students could potentially uncover trends that could inform relevant pedagogical considerations for DGBL. Directly applying gameplay behavioural strategies identified in the entertainment game industry to education will be illogical, as the differences between educational and entertainment DGs mean that differences in play behaviours are inevitable. One obvious reason for this is that entertainment games are voluntarily played for fun, while educational DGs are frequently obligatory. This particular fact affects the autonomy in choices made and overall play behaviours of students (Boyle *et al.*, 2011). Furthermore, gameplay behavioural strategies differ from one game to another, more so between learning games and entertainment games. Therefore, there is a need for research into gameplay behavioural patterns in educational DGs, an issue that the current chapter aims to address. The study in this chapter sets the stage for more research into the mapping of the gameplay behaviours of students to gain a better understanding of how these affect learning, teaching and assessment in educational DGs and other immersive learning environments. The contribution made by this study does not claim to be exhaustive, however, it presents a preliminary understanding of gameplay behaviours in an educational DG that could inform future studies. Furthermore, an assessment of gameplay performance is also carried out to explore the use of log data for assessments.

5.1.2 Game-based assessment using log data

Considering the calls for more authentic assessments (Villarroel *et al.*, 2019; McArthur, 2022), and the wider applications of DGs in education (Connolly *et al.*, 2012), a few studies have explored the use of game log data for assessing learning. For example, Loh *et al.* (2015) and Loh and Sheng (2015) mapped the action sequences performed by university students in a game. By calculating the Jaccard coefficients of the in-game actions, they were able to group students into expert and novice categories. The Jaccard coefficient is a similarity measure used to estimate the similarity between two data items (Loh and Sheng, 2013). Their use for assessments of performance in DGs is rare, probably because they require the initial identification of the ideal gameplay actions as a basis for ranking the performance of students. For most DGs, it is often difficult to specify a single ideal solution to game tasks given the many possible and valid solutions to any given task. This potentially limits the application of Jaccard similarity measures for game-based assessments.

Alternatively, cluster analyses have been used in a few studies to explore the gameplay performances of students (*e.g.* Hou, 2012; Kerr and Chung, 2012). In one study, cluster analysis was used to measure and identify key features of the performance of primary school pupils in a maths game (Kerr & Chung, 2012). Using the log data from an educational game called Save Patch, Kerr and Chung (2012) were able to determine the performance of students based on their gameplay strategies. In the study of Hou (2012), cluster analysis was also used to identify gameplay clusters using the log data of primary and high school students in a massively multiple online role-playing game (MMORPG). The study identified three clusters of gamers based on their gameplay activities: highest-participation gamers, high-participation gamers, and ordinary-participation gamers. Although cluster analysis was used by these studies to classify gameplay strategies and actions of students, it can also be useful for exploring patterns or similarities in gameplay data. This makes it practical for classifying students based on their performance in DGs. Cluster analysis can be a convenient statistical method for assessing performance in DGs where assessment components were not originally embedded during the game design. When assessment components are not incorporated in a DG, it is often challenging to determine the performance of students without analysing log files. These log files of the gameplay of students can be collected and analysed using clustering techniques to identify different performance groups and their characteristics.

The use of log data for performance and behavioural assessments in DGs highlights their relevance to educational research and informs the current study. Log data provides real-time process information on the actions of students and thus, can be very useful for measuring learning performance in game environments (Shute & Ke, 2012; Shute *et al.*, 2017). These can be valuable alternatives to traditional tests as significant correlations have been found between in-game performances of students and their performance on traditional multiple-choice tests (Sanchez *et al.*, 2022). The ability to store and retrieve real-time gameplay data that provide detailed action-based information about students offers invaluable benefits to research. Log data is utilised in this chapter because they provide useful information that is otherwise difficult to obtain with traditional assessments. These data can be used to infer the proficiencies of students based on their actions in the game rather than their responses to multiple-choice questions. The outcome of this analysis will serve as a backdrop for understanding the gameplay behavioural patterns of students in educational DG environments.

Only a handful of studies have reportedly used log data for assessment in engineering education disciplines, as highlighted in Chapter 2, with fewer studies exploring the gaming behaviours of adult learners in educational DGs. This chapter, therefore, presents an understanding of the gameplay behaviours of engineering students. It also presents the use of unstructured log data to measure engineering-relevant competencies based on the performance of students in a DG. The findings of this study would be particularly useful to educators and researchers in the field of game-based learning and immersive learning.

To accomplish these goals of this chapter, the following research questions will be answered in this study:

RQ1 What sequential behavioural patterns exist in the gameplay of engineering students?
RQ2 What performance clusters can be identified from the gameplay log data?
RQ3 What are the characteristics of the solutions in the performance clusters?

5.2 Research Methodology

This research uses a quantitative research method to answer the outlined research questions. Quantitative content analysis (QCA) was the research design adopted for this study. Increasingly used in educational technology research, QCA is characterised by an objective and systematic method for describing communication (Rourke & Anderson, 2004). QCA is defined as a research technique for making valid and replicable inferences from data (Krippendorff, 1989). A more commonly used definition of QCA was given by (Berelson, 1952) to be “a research technique for systematic, objective and quantitative description of the manifest content of communication” (p.18). Often used to tally the occurrence of events or contents, it can also be used to draw inferences about constructs (Rourke & Anderson, 2004). Depending on the study, data used for QCA can be text, audio, or any data that occurs in sufficient numbers and is meaningful to a given field (Krippendorff, 1989). QCA has been used in research for communication content analysis (Rourke *et al.*, 2001; Bullen, 1998). In this study, QCA procedure was used to answer the first research question by keeping count of the number of unique solutions used by the students to complete the game tasks. Similar to qualitative content analysis, there are some limitations and criticism of the validity of QCA as a research design. One of those concerns pertains to making inferences based on the outcomes of a QCA (Rourke & Anderson, 2004). Given the challenges with demonstrating the validity and reliability of data for QCA, Rourke and Anderson (2004) recommend using QCA to make descriptive rather than inferential conclusions. For this study, QCA is used to describe the behaviours of students in a game-based environment based on the number and characteristics of their solutions to the game tasks.

To answer the 2nd and 3rd research questions, cluster analysis was performed to identify performance clusters in the data. Cluster analysis is a classification technique used to group similar and homogenous sub-samples of cases or people as emphasised in Chapter 4 (Cohen *et al.*, 2017). Useful for discovering groups in data, it measures distance, proximity and similarity between data points. A hierarchical clustering technique was used to explore the number of groups or clusters in the data, while K-Means clustering was used to identify cluster characteristics, as detailed in Chapter 4.

5.2.1 Participants

A total of 58 chemical engineering students from KU Leuven and Imperial College London took part in this study. There were 39 male and 18 female students. 97% of the participants were between the ages of 20 and 29 years as shown in table 5.1. A purposeful sampling method (Creswell, 2011) was used to recruit these participants that were in their 2nd and 3rd year of engineering study. These cohorts of students have the basic engineering knowledge needed to solve the game tasks efficiently. Their interactions and solutions to the game tasks would help develop a detailed understanding of the gameplay behaviours of students with engineering backgrounds.

5.2.2 Data collection

This study serves as a follow-up to the study described in Chapter 4 and the data used for this study were the same data collected during the study presented in Chapter 4. Since the study was carried out during the peak of the COVID-19 pandemic, a 2-week window was given to students to play the CosmiClean game. This was done remotely at their convenience. Participants were encouraged to solve as many levels of the game as they could but were required to solve a minimum of 25 levels within the 2-week timeframe given. They were asked to solve the tasks as a chemical engineer would, paying attention to the sequencing and the configuration of the processors selected to ensure optimal solutions to the game tasks. An online survey and CosmiClean game log data were used for data collection. A questionnaire was used to collect demographic information of participants as well as their game experiences as presented in table 5.1.

Table 5.1: Demographic data of (n=58) engineering students who took part in the behavioural analysis study.

		Absolute Frequency
Gender	Male	39
	Female	18
	Unspecified	1
Age	Under 20 years old	2
	20-29 years old	56
Prior gameplay	Yes	54
	No	4
Gaming habits per week	Less than 5 hours	25
	5-10 hours	21
	11-20 hours	7
	21-30 hours	4
	31 hours or more	1

The CosmiClean recycling game was used for this study. Detailed in Chapter 4, CosmiClean is a product of the European Institute of Innovation & Technology – Knowledge and Innovation Community (EIT-KIC Raw Materials) game project of the European Union Horizon 2020 (EUH2020) in collaboration with LuGus Studio, Belgium. Its goal is to teach the heuristics of recycling processes to players.

5.2.3 Data analysis

To identify the behavioural patterns of students in the DG, a sequential pattern analysis of the gameplay activities of students was performed. First, log files were retrieved from the game after the 2-week window, then cleaned and processed in RStudio 4.0, and finally analysed using IBM SPSS version 27 software. Cleaning and processing the data involved extracting relevant information from the data collected and correctly formatting the dataset for analysis. The game metric used to explore the sequential behavioural patterns of engineering students in the DG is the number of unique solutions provided by students for each level of the game. Unique solutions are solutions that are different from the others in terms of the types and sequencing of processors used in the game. For performance assessment, the average energy used on each level and the number of resources used per level were analysed. A descriptive analysis of the gameplay metrics is presented in table 5.2.

Table 5.2: Descriptive results of gameplay activities of (n=58) engineering students in CosmiClean.

	Mean	**Std Dev**	**Maximum**	**Minimum**
Levels Completed	25.9	14.6	57	3
Energy Expended per Level completed	53.5	32.4	202.8	16
Unique Solutions per Level	6.2	4.4	26	1

The game has 57 levels and a total of 11 resources (2 non-processors and 9 processors) available to players. To solve each level, participants were required to drag and drop suitable resources to the recycling line. They were to choose from a catalogue of resources provided. These differ from one level to another. For each level solved by a student, the type and sequence of assembly of the resources were logged. To analyse these logged data, the string codes used were converted to numeric codes and are presented in table 5.3.

Table 5.3: Resources available in CosmiClean.

Resources	**Function**	**Symbol**	**Code**
Conveyor	Connects two resources and transports materials from one to the other	A	1
Receptor	Collects predefined material	N	2
Sieve	Separates materials by size	J	3
Melter	Separates materials by melting temperature	M	4
Magnet	Separate ferrous metals from other materials	I	5
Shredder	Separates materials by size reduction	H	6
Eddy Current Separator	Separated non-ferrous metals from other materials	E	7
Stream Separator	Separates materials based on the state of matter	L	8
Boiler	Separates materials by boiling temperature	G	9
Dissolver	Separates materials by solubility	D	10
Centrifuge	Separates materials by density	C	11

The resources outlined in table 5.3 are presented in their order of appearance in the game. For instance, in Level 1 of the game, players are required to connect the material feed to the receptor. For this level, only the conveyor is needed and provided. For Level 5, players are provided sieves and conveyors, and they are required to separate three distinct materials given their particle sizes. Conveyors, sieves, melters and magnets are provided in Level 12, and players are required to choose and sequence the most suitable resources to separate glass, iron, and bricks of similar particle sizes. The resources available to the players increase from 2 in the first level to 11 in the last levels of the game. Different solutions are possible at each level of the game as shown in table 5.4. The solution to Level 1 was the same for all students, 1-1-1-2, which means students placed the resources in the order: conveyor-conveyor-conveyor-receptor. For some other levels of the game, multiple unique solutions were used by the students.

Table 5.4: Examples of coded sequential solutions of some of the (n=58) engineering students.

Student ID	**Level 1**	**Level 3**	**Level 5**	**Level 7**	**Level 12**	**Level 14**
1	1-1-1-2	3-1-2-2	3-3-2-2-2	1-4-2-2	4-4-2-2-2	5-2-2
2	1-1-1-2	3-2-2	3-1-3-2-2-2	4-2-2	4-4-2-2-3	5-2-2
3	1-1-1-2	1-3-2-2	3-2-3-2-2	4-2-2	4-2-4-2-2	5-2-2
4	1-1-1-2	1-3-1-3-2-3-2-2	1-1-1-3-2-1-1-3-2-2	4-2-2	4-4-2-2-2	5-2-2
5	1-1-1-2	1-1-3-2-2	3-2-3-2-2	1-4-2-2	4-2-4-2-2	5-2-2
6	1-1-1-2	1-1-1-1-3-2-2	3-2-1-3-2-2	4-2-2	4-2-4-2-3	5-2-2
7	1-1-1-2	3-3-2-2	1-1-1-3-2-1-3-2-2	4-2-2	4-2-4-2-4	5-2-2

5.3 Results and Discussions

5.3.1 Gameplay behavioural patterns

To understand the sequential behavioural patterns of engineering students across all game levels, the coded sequential solutions for all levels were analysed, and the unique solutions were extracted. The number of unique solutions to the game tasks was computed and is graphically represented in figure 5.1. High variabilities were found in the number of unique sequential solutions across all 57 levels of the game. This means that students used different sequences of solutions and varying numbers of resources for solving each level of the game.

More unique solutions to individual levels were seen in the first few levels of the game compared to the latter levels. Given the ease of the tasks in the first levels, it was interesting to observe high variability in the solutions used by students. DGs are designed to be engaging and to facilitate flow experience. Flow state, the complete absorption or optimal experience in an activity, hinges on the balance between challenge and skills, considered by Kiili (2005) as the "zone of proximal development".

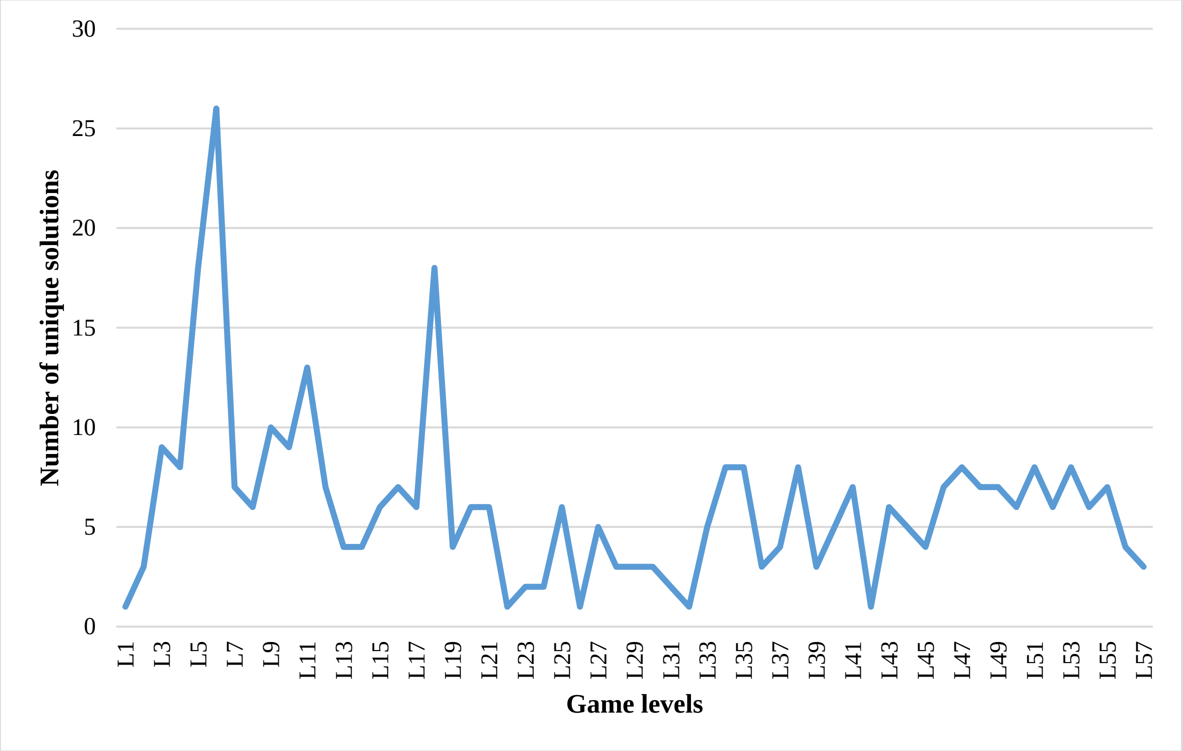


Figure 5.1: Sequential behavioural pattern of (n=58) engineering students across CosmiClean game levels.

DGs such as CosmiClean, designed with tasks that are incrementally challenging, are intended to support players through the gameplay, from presenting simpler tasks at the beginning to more complex tasks subsequently. The sequential solution patterns that inform the gameplay behaviours of students depict a trend from *exploratory* to more *refined or strategic* behaviours. An explanation for the high variability in the first few levels could be that at the start of the game, students were exploring all the options and testing out game elements and resources available to them as they tried to familiarise themselves with the game. This exploratory behaviour could have resulted in the high number of solutions used to solve less complex tasks at the start of the game.

As shown in table 5.4, the solution patterns of the participants were erratic at the first few levels with the use of excess conveyor belts and in some cases, more processors than necessary. For instance, in Level 3, the game task was to separate a mix of plastics and wood particles of different sizes. The optimal solution would be to use a sieve and two receptors. The mesh size of the sieve would need to be set to accurately separate both particles. From table 5.4, only student 2 can be seen to use this optimal solution while others used more resources to achieve the same goal. This trend can be seen to diminish from the 7th level. In level 7, students were required to separate moulds of a mix of iron and concrete given their melting temperatures. For this separation, the ideal solution would be to use a melter configured to melt the iron, and two receptors to collect each of the separated particles. The solutions of most of the students at this level were homogenous with only a few students using conveyors which were unnecessary.

Similar trends seen in this study were also identified in the study of Kang *et al.* (2017). Kang and colleagues reported several exploration activities of students during the first stages of their gameplay, indicating that the trend observed in the current study is not isolated. Considering that students were not given any prior training on the game mechanics, it is logical that they explore how the game works and thus attempt different valid routes to solving the game tasks. As students progressed through the levels, the exploratory patterns diminished despite the increased complexity of the tasks at those levels. Less erratic and more refined solutions were observed in these levels indicating that more "strategic" approaches were used by students to solve the given tasks. As students progressed through the levels, they would have become confident in themselves and would think through their solutions, leading to more optimised and fewer unique solutions.

The log data from the gameplay of students in this study paints a picture of their behaviours in the DG. The exploratory to strategic gameplay behaviour observed in the data explains the potential behaviours of engineering students in immersive learning environments, particularly one that is new to students. Based on these findings, it can be said that the sequential behaviour pattern of engineering students in educational DGs follows a trend from exploratory to more strategic behaviours. A different trend could emerge with a different cohort of students, a different game genre or when different game metrics are analysed. Nonetheless, this finding is valuable as it could potentially inform pedagogy, game designs and research.

5.3.2 *Performance clusters*

A 2-step cluster analysis involving hierarchical and K-Means clustering was used to partition students into groups based on their performance on game tasks. Since most students completed 25 levels of the game, the data of a total of 32 students who completed the first 25 levels were selected in order to make the analysis more consistent. To determine the optimal number of groups of students based on their gameplay performance, a hierarchical cluster analysis was first performed using the game log data. The variables considered here to measure performance were the average number of resources used, and the average energy expended in solving the tasks. The average number of resources implies the average number of resources such as conveyor belts, sieves and receptors used for the first 25 levels of gameplay, while the energy expended is the amount of energy consumed to complete the tasks. Formann's sample size recommendation (Formann (1984) cited in Dolnicar *et al.*, 2016) calculated in equation 5.1 indicates that the minimum sample size recommended for this analysis is 20. With data from 32 students, the minimum sample size requirement is met.

$$n_{\min} = 5 \times 2^{\chi} \qquad \textit{Equation 5.1}$$

$$5 \times 2^{2} = 20$$

where $n_{\min}$ is the minimum sample size; χ is the number of variables.

To begin the analysis, the scores on both variables were first standardised to make the relative weight of each variable equal. Ward's method was used for this as it has been found to be a more effective fusion method for hierarchical clustering compared to other methods (Landau *et al.*, 2011). From the dendrogram in figure 5.2, the optimal number of clusters in the data was found to be three, corresponding to the longest branches (length of horizontal lines in the dendrogram) that indicate unique clusters or groups. For the next stage, a 3-cluster K-Means analysis was performed. These analyses were carried out using IBM SPSS version 27.

From the results of the K-Means clustering, it was found that Cluster 1 consists of students who used the highest number of resources to solve all 25 levels and expended less energy to do so compared to the other two clusters. Cluster 2 on the other hand consists of students who used the least number of resources and the highest amount of energy. The 3rd cluster consists of students who used the least amount of energy and fewer resources in their solutions. As shown in table 5.5, Clusters 1, 2 and 3 were labelled "Optimal Energy", "Optimal Resources", and "Optimal Solution" groups, correspondingly. To visualise the differences in performance of the three groups, the z-scores of the metrics used for the analysis, that is, the average energy and resources used per level were plotted and are presented in figure 5.3. By standardising the values of both variables to make their means and standard deviations 0 and 1 (standardisation), respectively, this figure provides a visual representation of the relative performance of the groups.

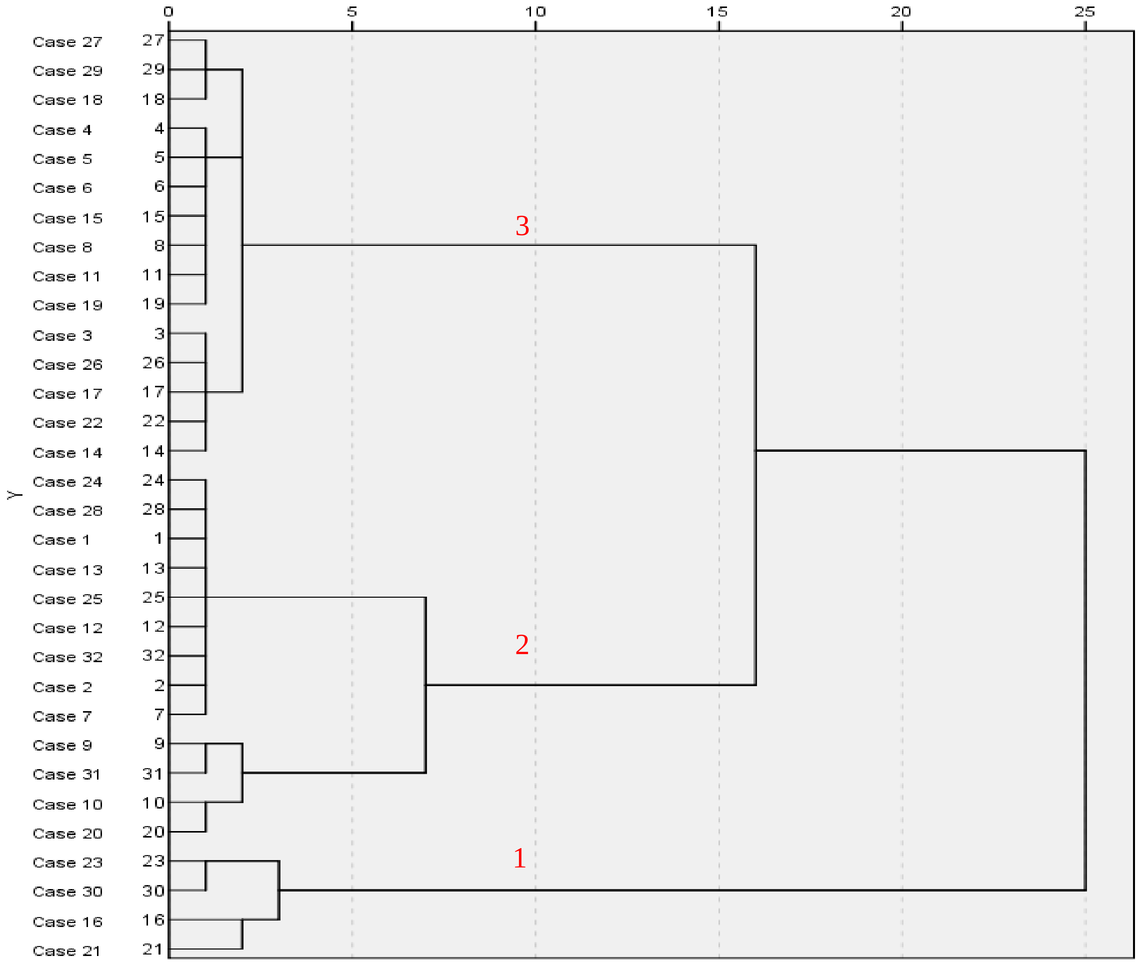


Figure 5.2: Dendrogram of hierarchical clustering using Ward's method with (n=32) engineering students based on gameplay data from CosmiClean.

Table 5.5: Cluster characteristics of the performance of (n=32) engineering students in CosmiClean.

	Cluster 1 **Optimal Energy**		**Cluster 2** **Optimal Resources**		**Cluster 3** **Optimal Solution**	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev

Average energy	49.6	9.7	107.1	17.9	46	10
Average resources	5.09	0.15	4.55	0.15	4.65	0.12
	N=5		N=4		N=23	

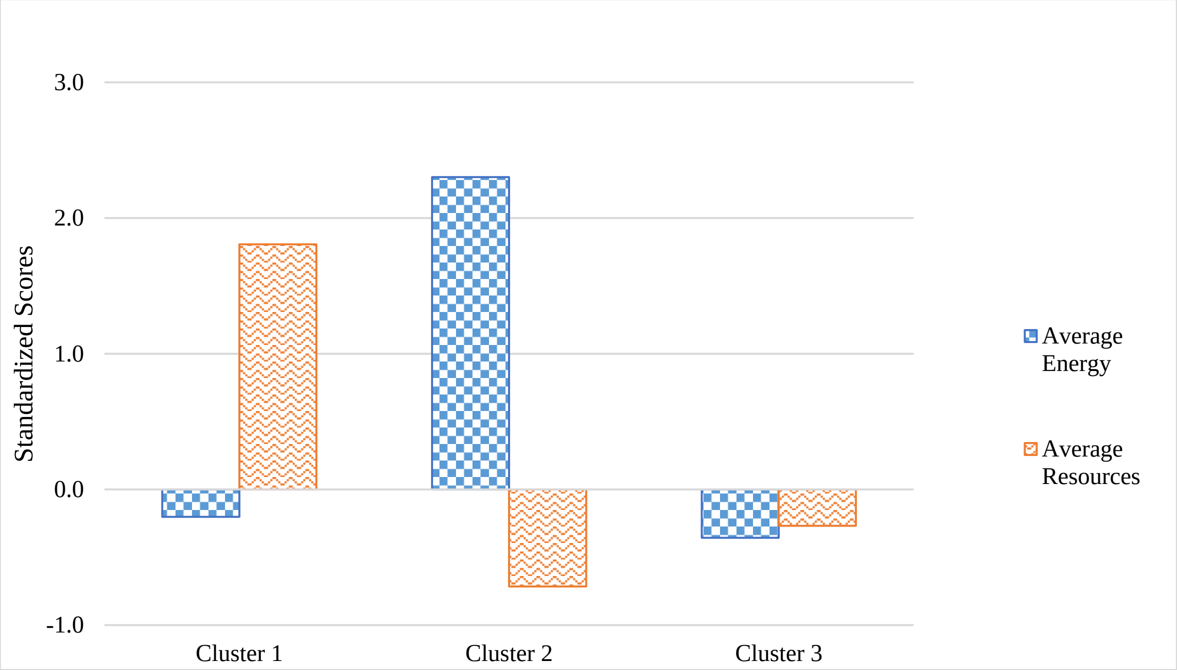


Figure 5.3: Visualisation of cluster performance of (n=32) engineering students in CosmiClean.

Of the three performance clusters identified in the data, the Optimal Solution group performed considerably better. Their solutions are considered the most sustainable of the three, a factor that is of utmost importance in the chemical engineering discipline (Glassey & Haile, 2012). The solutions of the students in this group involved the use of fewer resources, fewer than those of the Optimal Resource group but similar to those of the Optimal Energy group as shown in table 5.5. This implies that their solutions were thought through, consisting of mainly relevant resources and the use of optimal sequences. Additionally, the students in this group considered the energy requirements of the resources used in their solutions. They prioritised energy-efficient resources over alternatives where possible. They also seemed to have paid attention to the configuration of their processors to achieve energy efficiency. With most of the processors configurable, students had the option to make changes to the configurations to suit the intended separation operation, which would lead to more energy-efficient solutions. For instance, to separate a salt solution, students could configure a boiler to 100 ^{o}C which is energy efficient for the required task (boiling temperature of water is 100 ^{o}C) or leave it at the default setting of 200 ^{o}C which will still separate the solution but with higher energy consumption. Interestingly, 72% of the solutions of the students fall into the Optimal Solution group which is a good indication that the majority of the students viewed the game tasks from an engineering point of view and thought through their solutions as expected.

The performance of the other two clusters of students is considered weaker than those of the Optimal Solution group. However, only 28% of the students fall into these categories. Nonetheless, the solutions of the Optimal Energy group were quite similar to those of the Optimal Solution group discussed above. As shown in table 5.5, students in this group expended far less energy compared to the Optimal Resource group, but a little more than the Optimal Solution group. When compared to the other two groups, the Optimal Energy students used the most resources when solving the given tasks. This indicates that although these students may have configured and sequenced their processors appropriately to achieve energy efficiency, they used far more resources than was necessary. The Optimal Resource group, on the other hand, used the least number of resources compared to the other two groups, but their solutions consumed higher amounts of energy. This indicates that students in this group used fewer resources to achieve the necessary separations, but they probably used the least energy-efficient processors and also failed to configure their processors to achieve energy efficiency. Overall, the solution patterns of the Optimal Resource group could be considered the worst of all three solutions obtained. Process optimisation is a fundamental requirement in chemical engineering operations that demands the selection of the most efficient process unit options that offer economically optimum performance and operations from possible alternatives (Chaves *et al.*, 2016). The performance of the Optimal Resource group shows that process optimisation considerations were not made while solving the game tasks. Considering that students were specifically instructed to solve the game tasks as a chemical engineer would by using only optimal solutions, the performance of this group of students suggests a lack of understanding of process optimisation or indifference towards the instructions given.

5.3.3 Solutions characteristics

To better understand the three performance groups obtained from the gameplay data, the mean scores on the considered metrics were plotted against the cluster groups. First, a look at the trend of energy consumption across all 25 levels presented in figure 5.4 shows the similarity between the Optimal Energy group and the Optimal Solution group. It also shows that overall, the Optimal Resource group did not deviate systematically, but consumed significantly more energy in some game levels, with a significant spike at game level 20.

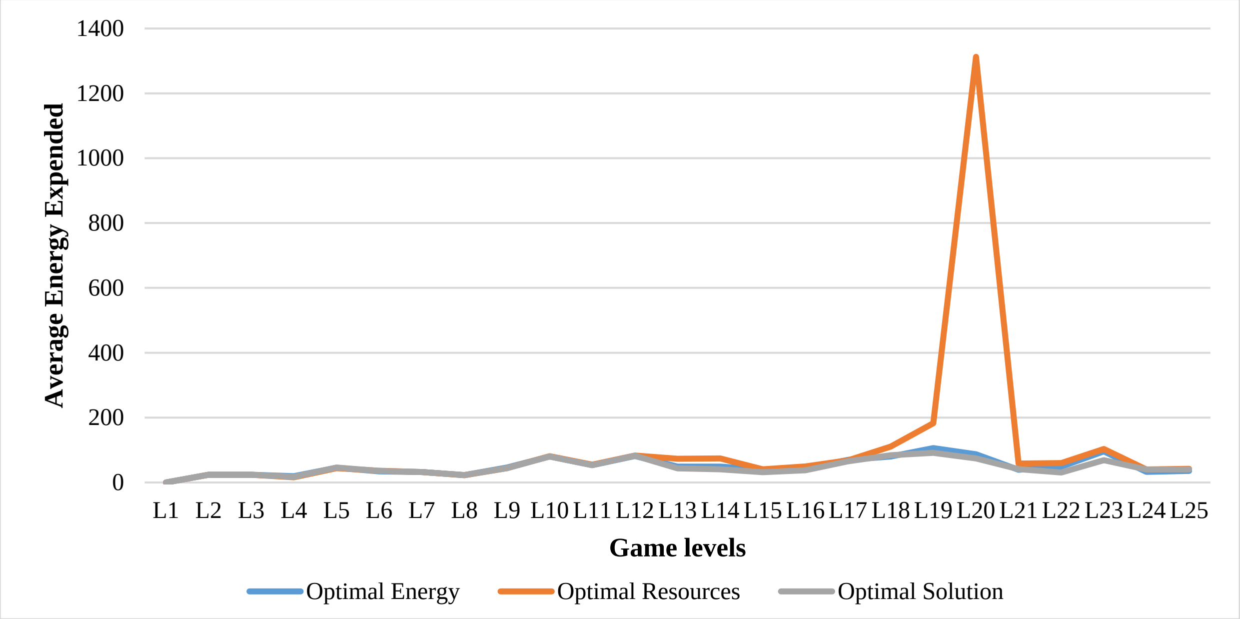


Figure 5.4: Visualisation of energy consumption of cluster groups of (n=32) engineering students across the 25 levels of CosmiClean.

The solutions of all four students in this group were found to have consumed an average of over 1200 units of energy on this level compared to an average of about 90 units used by the other groups. The game task for this level required students to separate 65 units of iron and a mould containing a mixture of one unit of iron and four units of bricks. The solution used by most of the students employed a sieve to separate the 65 units of iron first, then a shredder to shred the mould of iron and bricks into single units, and finally, a magnet separator to separate the iron from the bricks. On the other hand, the solution used by all the students in the Optimal Resource group involved a shredder first to shred all the materials before passing these through a magnet separator. This is an energy-intensive process because the energy consumed by a shredder and a magnet separator to process one unit of material is higher than that required by a sieve. With all the materials for separation passed through the shredder and the magnet, much more energy is consumed than when a sieve is first used to separate a bulk of the materials. This highlights the limitation of the solutions used by the Optimal Resource group. It also shows that the basic rules of separation operations which recommend the separation of easily separable materials first to minimise the streams to be handled by other units subsequently (Seader *et al.*, 2011), were ignored by these students. These findings demonstrate that log data and cluster analysis can provide detailed performance-based insights about students which can help educators assess specific learning outcomes and provide tailored feedback to the students.

Lastly, figure 5.5 provides a visual representation of the number of resources used by the different groups across the 25 levels. As initially seen in table 5.5, the trend shows that the Optimal Energy group used more resources on average across the levels compared to the other two groups.

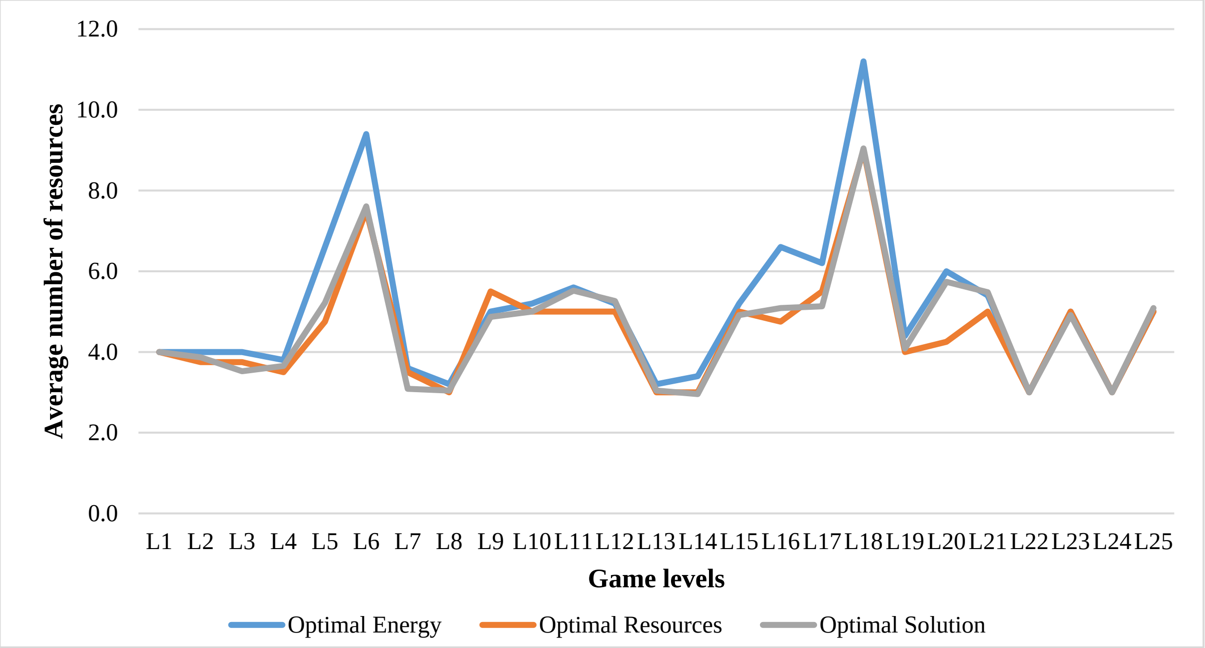


Figure 5.5: Visualisation of resources used by cluster groups of (n=32) engineering students across the 25 levels of CosmiClean.

Spikes in the number of resources used by all three groups were seen in Levels 6 and 18. Generally, the solution sequences used by the students showed that they used much more conveyors and sieves than were required for solving those levels. Although the energy expended from the use of conveyors and sieves is negligible in this game, in reality, this may not always be the case, as having unnecessary process units in place does have significant implications in terms of possible inefficiencies/losses as well as space and capital costs.

5.4 Implications of Findings

5.4.1 Assessment

It is common to find assessments administered to students after a few hours of gameplay in order to ascertain the effectiveness of the game. With external assessment methods mainly used for this, the outcomes have been interpreted favourably as detailed in Chapter 2. The outcome of the current behavioural analysis questions the validity of assessments administered to students after limited interactions with a DG. Given the findings that the initial behaviour of engineering students in a DG is exploratory, assessing and judging performance based on solutions to the first few levels or in other cases, the first few hours of gameplay, would neither be fair to the students nor lead to a valid interpretation of the competencies of students. When designing assessments to determine the learning effectiveness or the performance of students in immersive learning environments, the recommendation would be to allow ample time for students to familiarise themselves with the environments. This would ensure that performance in the learning environment is judged based on the strategic gameplay phase of students. However, the outcome of the assessment of students after extensive gameplay time may indicate mastery of procedural skills but may not necessarily mean a deep-level conceptual understanding of the subject taught, as was found in a study by Crawford and Colt (2004). Where considerable gameplay time is allowed and conceptual understanding is of interest, it might be worth exploring multimodal assessment methods to compare in-game performance with performance on a different type of assessment. If limited time is available for an intervention study, pre-game training can be offered to students to shorten the exploratory phase of gameplay. Students can be given appropriate support to enable them to navigate the initial levels quickly and accurately. Doing so would likely reduce the exploratory phase of gameplay to allow assessments to be carried out sooner. There is however the risk of impacting independent learning, and discouraging active participation and experimentation that enhances deeper learning. An appropriate balance should be struck to ensure that tutorials do not inhibit curiosity and that the educational levels of the students are taken into consideration. Taking into account this gameplay behaviour from exploratory to strategic will likely reduce the effects of prior game experiences on performance. It will also increase validity, reliability and fairness in game-based assessments, and improve the confidence of students in game-based assessments.

5.4.2 Research

This study demonstrates that more research into DGBL is needed. The observations made from this study challenge the current understanding of how students learn in DGs. Based on the findings from this study, further research into the behaviours of adult learners in different disciplines is needed to determine if the patterns found in the current study are confirmed in a wider context. Learning theories relevant to immersive learning technologies are needed to enhance the understanding of learning in these environments and to improve their effectiveness for education. Also, the findings about the gameplay behaviours of students should be explored further through interviews or focus group discussions to understand the underlying factors in the observed trend. Based on the findings of this study, the use of log data for research in immersive learning should be considered alongside other research instruments currently used.

5.5 Limitations

One limitation of this study is the small sample size of participants of the study. This limitation means that the outcomes of this research cannot be generalised to the entire population and may differ with a different group of students or a different immersive environment. Another limitation of this study is the subjectivity of data analysis and interpretation. The data analysed were gameplay data available to the author at the time of the study. The metrics considered for analysis and the interpretations of the results were based on the relevance placed on them by the author.

5.6 Conclusions

The outcome of the behaviour and performance analyses of engineering students in a DG shows that students exhibit different behavioural patterns across gameplay. From uncovering exploratory and strategic behaviour patterns to identifying three performance clusters and their characteristics, log data provided detailed information for understanding the interactions of students in the game environment. It would have been useful to understand the reason for the identified pattern by asking participants why they made certain gameplay decisions. As it was not possible to carry out this additional data collection at the time because students failed to sign up for the study, future studies should consider repeating this research and triangulating observations with findings from interviews or discussions with participants. Nonetheless, the outcome of this study should encourage more extensive use of log data for assessments when using DGs. The visual representations of solution characteristics obtained from game log data provide clearer views of the gameplay actions of students. The data highlight the strengths and weaknesses of each student and demonstrate the effectiveness of log data of gameplay for understanding performance, designing authentic assessments and providing relevant and tailored feedback to address fundamental misconceptions. The exploration of log data for behavioural analysis should also inspire research into understanding gameplay behavioural implications for DGBL. To fully exploit the potential of immersive learning technologies for education, an in-depth understanding of the learning and behavioural psychology relevant to these non-traditional learning environments is crucial. The type and amount of data generated during gameplay as well as the ease of storing and retrieving these data make log data invaluable resources that can be used to expand understanding of the pedagogical implications of immersive learning environments.

5.7 Note: Part of this study is published in the **Gamification and Social Networks in Education** book: Udeozor, C., Russo Abegão, F., & Glassey, J. (2022). Exploring Log Data for Behaviour and Solution Pattern Analyses in a Serious Game. In Bakan, U & Berkeley, S. (Eds), *Gamification and Social Networks in Education*. MacroWorld Pub. Ltd, Bayrakli-İzmir**.**

Chapter 6. A Game-Based Assessment Framework for Immersive Learning Environments

In the previous chapters, the focus was to gain an understanding of the applications of digital games (DGs), virtual reality (VR) and augmented reality (AR) in the engineering education context. The key interest was to identify the assessment implications of the use of these technologies in order to inform the design of a robust assessment framework. The outcomes of these chapters provided valuable information which would enable the design of an effective assessment framework, as well as assessment guidelines that would be relevant to educators in higher education (HE) institutions. The current chapter utilises a design-based research methodology to present the development of an assessment framework grounded in the Evidence-Centred Design conceptual framework and the principles of Constructive Alignment. The framework is also a product of the experiences gained from the studies in the previous chapter. It is designed to take into account the challenges encountered while measuring the performance of students in the previous chapters. The framework outlines steps and components necessary for the effective design assessments of learning in DGs, VR or AR environments. This chapter also demonstrates the application of the framework with the design of a game-based assessment for CosmiClean. Subsequently, the outcome of the evaluation of the assessment frameworks carried out by the author, an independent educator and the supervisors of this research project is described. The observations and feedback from the evaluation indicated the need for a simplified version of the framework which will be presented in the next chapter.

6.1 Introduction

Assessment of learning is at the core of formal education and is a crucial part of the learning process of students. Assessments provide opportunities to measure the progress made by learners on the learning objectives, evaluate teaching strategies, and pass judgement on the abilities of students based on performance in assessment tasks. They are considered tools for determining the level of knowledge, skills or abilities of students based on what they say, do or make in a given setting (Mislevy, Steinberg, *et al.*, 2003). In HE, established guidelines are often followed when designing classroom learning activities, including assessments of learning. Some of these guidelines include the SOLO taxonomy (Biggs & Collis, 1982), Bloom’s taxonomy (Anderson, 2013) and Constructive Alignment principles (CA) (Biggs & Tang, 2010). These guidelines serve to provide logical approaches for designing effective classroom practices.

Assessments are very important activities that could have life-changing implications for students. The chance to get accepted into a university programme of choice and subsequently pursue their dream careers depends primarily on the outcomes of high-stake exams such as A-Levels. For HE students, the likelihood of being shortlisted for competitive entry-level graduate roles is often influenced by their grades. With the changing demands of the workplace and the need to improve learning effectiveness, advanced technological tools are finding their way into HE pedagogy, stretching the boundaries of existing pedagogical strategies.

The distinct nature of emerging technologies like VR and AR, when compared to other traditional teaching and assessment tools, raises concerns about the validity of existing pedagogical practices for immersive learning. The interest and growing adoption of immersive technologies in HE demonstrate the need for appropriate pedagogical practices tailored to these unconventional learning environments.

The design of assessments in HE often follows the principle of Constructive Alignment which requires an adequate connection between learning objectives, learning and teaching (L&T) activities and assessment tasks (Biggs & Tang, 2010). Learning objectives are the expected results, intended outcomes, or intended changes that a teaching activity aims to deliver (Anderson, 2013). Assessments can be aimed at monitoring progress made by students in order to provide feedback and improve teaching strategies in the case of formative assessments, or for the grading and validation of knowledge in summative assessments. In either case, assessments should be well thought-through and designed to ensure accurate measurement of the learning objectives of interest with the chosen assessment method.

Traditionally, assessments are administered as open/closed book exams, in the forms of multiple-choice tests, short answer questions, essays, reports or portfolios designed to measure specific learning objectives. However, the concerns about the validity of the use of some of these assessment methods in HE have led to conversations about the use of more authentic assessment methods that measure higher-order level cognitive processes (Villarr*o*el *et al.*, 2019).

Assessments are thought to be authentic when they are based on the activities of students that replicate real-world work or tasks (Svinicki, 2004; McArthur, 2022). According to Ashford-Rowe, Herrington, and Brown (2014), authentic assessments should be challenging, performance or product-outcome based, must ensure knowledge transfer, enhance self-reflection and self-assessment, be contextual and accurate, and should encourage discussions and collaboration. These elements considered critical for authentic assessments are not always easy to apply to classroom assessments but can be achieved through problem-based learning and assessments (Merrett, 2022).

The advances in technology and the relative affordability of immersive technologies have led to a rise in their use for teaching, training and for authentic assessments in HE and professional settings. Serious DGs are now increasingly being used for graduate recruitment with large multinationals like McKinsey, Shell, Unilever and Deloitte incorporating these into graduate-level recruitment processes (Bina *et al.*, 2021; Kashive *et al.*, 2022). For engineering education, there are considerable numbers of studies reporting on the use of DGs for teaching and assessment as reviewed in Chapter 2. In addition to providing an active learning environment, VR, AR and DGs offer complex learning environments that are challenging, problem-based, and realistic, offering students the opportunity to apply knowledge and skills to different real-world contexts while supporting collaboration and self-assessment. All of these elements considered critical for authentic assessment are achievable with immersive technologies.

The interactions of students within immersive environments generate large amounts of cognitive and non-cognitive data that are captured in log files which provide an estimate of their knowledge and skills (Shute *et al.*, 2017). The ability of immersive technologies to simulate realistic environments provides learners with accurate representations of world-of-work environments and a sense of the real-world applications of the knowledge learned in classrooms. These technologies also allow educators to assess students based on how they apply multiple cognitive skills to complex tasks. Importantly, such assessment is based on what *students do* (performance-based) rather than what *they say*. Advocates of assessments with immersive technologies believe that there is a lot of promise with these tools for assessment (DiCerbo, 2017; V. Shute *et al.*, 2017). Immersive technologies enable the assessment of skills and constructs that are otherwise difficult to measure with traditional assessment items. They are also thought to reduce test anxiety and offer more valid and reliable assessments of competencies (DiCerbo, 2017; V. Shute *et al.*, 2017).

One of the major obstacles found to be faced by educators who consider the use of immersive technologies for formal classroom education is the assessments of learning in these environments (Razak *et al.*, 2012; de Freitas, 2006; Routledge, 2009). Educators lack clarity on how to measure learning and make valid inferences about what students know and can do in immersive environments. In Chapter 2 of this thesis, the two methods of assessments often used for immersive learning environments were discussed in detail (Ifenthaler, Eseryel, and Ge, 2012; Shute, Ke, *et al.*, 2017). Embedded assessments that either provide generic summarised performance-based details such as completed tasks, tasks completion times and points accrued, or other detailed performance-based process data of students are generally recommended (Shute *et al.*, 2018). However, external assessment methods, which are often multiple-choice tests or other forms of assessments external to immersive environments, are more prevalent in engineering education. The preference for external assessments is attributed to the complexity and resource-intensive process of designing embedded assessments. Analysing unstructured process data from log files containing data on the actions of students can also be laborious (Loh, 2009; Wallner & Kriglstein, 2012). Designing assessments around games or simulation tasks can be daunting for educators. The challenge with designing embedded assessments is further complicated by the busy schedules of educators whose workload limits their exploration of new learning technologies.

Attempts have been made to develop assessment frameworks for designing embedded assessments for immersive learning environments. Loh (2012) proposed the Information Trails system for game-based assessment. Information Trails emphasise the recording, storage and retrieval of relevant gameplay process data for analysis. The system was proposed over a decade ago to demonstrate the potential of gameplay data for assessments of learning following the growth of the entertainment games industry. Information Trails system is simply what it says, the collection and analysis of process data. It does not offer any information about the design of assessments nor the identification of relevant game tasks for the assessment of any intended learning outcomes. This could be the reason why reports on the application of Information Trails for assessment are very sparse in the literature.

One other assessment framework which has been applied to game-based assessments in relatively more studies is the Evidence Centred Design (ECD) by Almond, Steinberg and Mislevy (2003). ECD provides an elaborate framework for assessment designs for a wide range of environments, including computerised standardised tests and simulation-based tests (Mislevy, Almond, *et al.*, 2003). For the assessments of learning in DGs, the ECD framework has been mostly used by games developers and researchers for the design of embedded assessments (see Kerr and Chung, 2012; Almond *et al.*, 2014; Jaffal and Wloka, 2015; Shute, Rahimi and Emihovich, 2017). Although effective for the design of assessments for immersive learning environments, the design process is often complex and time-consuming and requires advanced statistical and machine learning skills, limiting its applications to skilled developers and researchers (Westera, 2019; Westera *et al.*, 2020; Kim *et al.*, 2016). It is also most effective for use during the design phase of the immersive learning environment, making it less useful for assessment designs around pre-existing immersive environments. These shortcomings make the use of the ECD framework particularly challenging for educators who lack both the time and resources to develop new immersive tools.

For wider adoption and broader impact of immersive learning technologies in HE, the procedure for the assessment of learning must be clear, simple and educator-friendly. A robust assessment framework that can be applied to the design of both external and embedded assessments would certainly promote the use of immersive learning technologies for performance assessments. Any assessment framework targeted at educators must be relevant to the design of external assessments. This implies that, in addition to being useful for the design of embedded assessments, such frameworks should offer guidance on the design and implementation of familiar external assessment types for the measurement of learning in immersive environments. While discouraged for assessments when using immersive learning technologies for L&T, external assessments are still the preferred choice as evidenced in Chapter 2 of this thesis. In this chapter, therefore, an assessment framework developed for immersive learning environments is introduced. It is designed to be educator-friendly and effective for designing both external and embedded assessments. To demonstrate the application of the framework, this chapter also presents the design of an embedded digital game-based assessment. The outcome of the evaluation of the assessment framework is also presented.

6.2 Research Methodology

The design-based research (DBR) methodology is adopted in this chapter to develop an assessment framework for measuring the performance of students when using immersive learning technologies. DBR is an applied educational research methodology used to provide practical solutions to educational problems. Wang and Hannafin (2005) defined DBR as a systematic, yet flexible methodology, intended for enhancing educational practices. DBR is application focused and uses either qualitative, quantitative, or mixed-methods data to achieve its goals. The qualities of the DBR methodology include its focus on real educational context, the design and testing of interventions, multiple iterations and collaborations among researchers and educators (Wang & Hannafin, 2005; Anderson & Shattuck, 2012). It is only common among educators and educator-researchers as a methodology for identifying and solving practical classroom issues. DBR methodology is increasingly used in educational research because of its quality and unique applications to solving practical issues, but there are criticisms concerning its use as a research methodology. The criticisms include subjectivity and the lack of rigour found in other research methodologies (Anderson & Shattuck, 2012; Creswell, 2011). Given that this research design is often used within the school context with educators who are not necessarily skilled researchers conducting the research, the concerns about how much rigour they apply come into question. Researcher biases inherent in qualitative research can limit the validity of the outcomes of a DBR, however, it is argued that these can be useful for research (Anderson and Shattuck, 2012). It is believed that the biases, insights and deep understanding of the context that these educator-researchers have, allow them to better interpret the research data obtained.

Regardless of these criticisms, DBR is considered valuable to education. For design research, such as that presented in this chapter, DBR is considered most appropriate given its emphasis on the design, development, iterations and evaluation of a practical solution to an educational problem. The lack of rigour and systematic procedure limitations of DBR is unlikely to affect the quality of the research presented in this chapter. This is because the assessment framework introduced herein is underpinned by established conceptual and theoretical models and by the findings from the preceding chapters.

To provide structure to the research presented in this chapter, a systematic process outlined in figure 6.1 is followed to achieve the goal of developing an assessment framework. The research problem and the goal of the research are first outlined, followed by the theoretical considerations for achieving the set goal. Next, the assessment framework is introduced and an evaluation of the framework is subsequently presented.

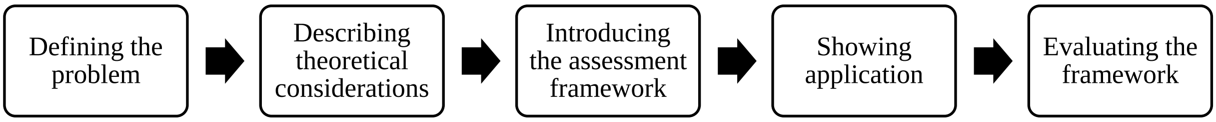


Figure 6.1: Systematic process for the development of an assessment framework for designing assessments for immersive learning.

6.3 Defining the Problem

Immersive learning technologies offer many benefits to HE. One of the opportunities that these tools offer is the ability to assess difficult-to-measure constructs and higher-order cognitive abilities processes (Voorhis & Paris, 2019; Rice, 2007). However, designing and implementing assessments in immersive learning environments are not common practices in HE. A reason for this is that educators often do not know how to design assessments to measure learning in immersive environments. The goal of this research is therefore to develop a robust and user-friendly assessment framework that can be used by educators for the design of VR, AR or DG-based assessments.

6.4 Theoretical Considerations

The Game-Based Assessment Framework (GBAF) introduced in this chapter is underpinned by two established conceptual frameworks: the Evidence Centred Design (ECD) framework and the principle of Constructive Alignment (CA).

6.4.1 Evidence-Centred Design Framework (ECD)

The ECD is a framework for designing assessments based on evidentiary reasoning. It is well known for ensuring the validity of evidence collected for assessments, and for its suitability for measuring complex competencies (Arieli-Attali *et al.*, 2019). The ECD enables the linking of the assessed competencies with assessment tasks (Mislevy, Steinberg, *et al.*, 2003). It is a product of the Educational Testing Services (ETS), the organisation known for developing, administering and scoring standardised tests such as the Test of English as a Foreign Language (TOEFL) and Graduate Record Examination (GRE). This framework was developed to enable the design of a broad range of assessment types, from standardised tests to portfolios and simulation-based tests (Mislevy, Almond, *et al.*, 2003). The ECD consists of models that specify the operational elements of an assessment and their interdependencies. As shown in figure 6.2, the ECD is made up of the Student Models, the Evidence Models, Task Models, Assembly Model and Presentation Model.

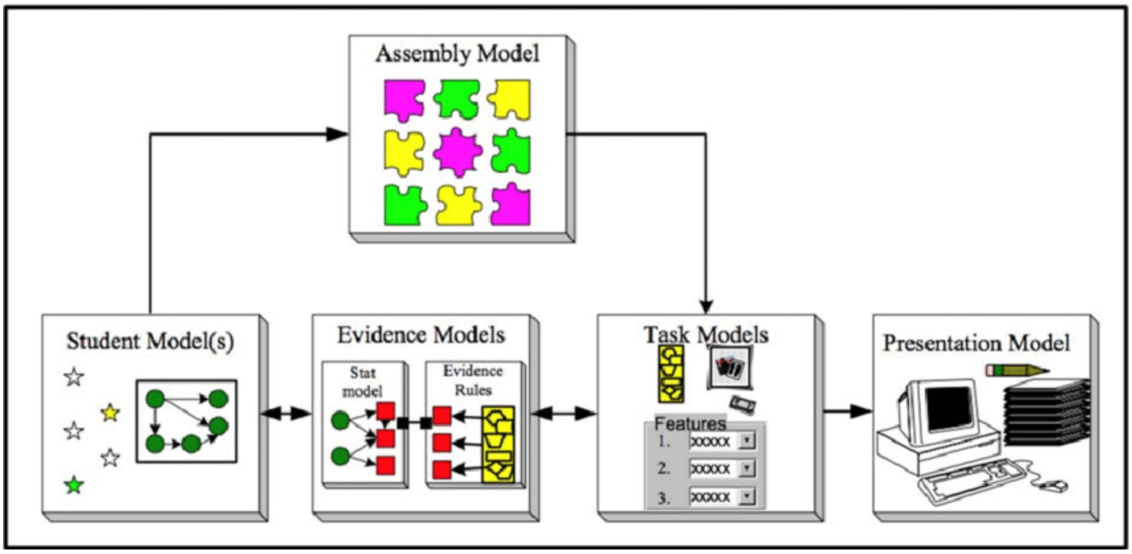


Figure 6.2: The ECD Framework (Mislevy, Almond, *et al.*, 2003).

Student Models: the student models answer the question, "*what are we measuring*?" These are sometimes also referred to as the competency or proficiency models (Shute & Ventura, 2013; Behrens *et al.*, 2012). The student model variables of the ECD define the skills, knowledge and abilities being measured. For instance, "apply heat and mass balances to non-reacting systems" could be a competency variable for measuring the knowledge of material and energy balance of 1st-year chemical engineering students. Knowledge of the competencies of students on this variable is initially unknown and is updated at every point in time during interactions with the immersive environment. Each value or outcome of the measured competencies is expressed by a probability distribution (Mislevy, Almond, *et al.*, 2003). In the case of multidimensional student models, Bayesian networks provide a graphical language for showing multidimensional associations (Almond, 2015).

Evidence Models: these models answer the question, "*how do we measure it*?" They describe how to update information about the student model variables based on evidence produced by students in a given task (Mislevy, Almond, *et al.*, 2003). They provide evidence of the competencies of students by linking what they do to the measured competencies. An evidence model is made up of two parts: evidence rules and measurement models. Evidence rules describe how the performance of a student in a given task is summarised from observable variables. In a standardised test, evidence rules guide the response scoring procedure. The measurement model on the other hand provides details of the relationships between the student model variables and the observable variables. It contains statistical models for the accumulation and synthesis of evidence across tasks and thus, guides the summary scoring procedure (Mislevy, Almond, *et al.*, 2003). Bayesian Inference Networks are the preferred statistical approach used by many scholars due to their graphical underpinning that aligns well with the principles of ECD (Behrens *et al.*, 2012).

Task Models: the task models describe a group of tasks that are presented to students to assess proficiency in a given subject. The task model answers the question, "w*here do we measure it*?". Groups of tasks or activities in the task model elicit observable evidence of unobservable competencies in the student model (Shute & Ventura, 2013). In each group of tasks, there are typically several tasks measuring the same variable (Mislevy, Almond, *et al.*, 2003). In a standardised test, for instance, each measured competence will generally require different task models and different sets of items or questions would be needed to assess them.

Assembly Model: this model answers the question, "*how much do we need to measure*?" It describes how much evidence or how many tasks are needed to make valid inferences about the students (Almond, 2015). It also ensures that multiple possible forms of tasks presented to students are comparable, especially in computer adaptive testing where students receive unique test forms (Mislevy *et al.*, 2012). For a scoring engine, assessment designers must construct a mathematical realisation of the student model and an evidence model for each task option (Kim *et al.*, 2016).

Presentation Model: This describes how the assessment tasks are presented to students. It provides specifications for how the other models are initiated in the delivery system (Mislevy, Steinberg, *et al.*, 2003).

The applications of the ECD framework to game-based assessment designs have largely focused on the first three models with less emphasis on the assembly and presentation models. The ECD framework has been used to design unobtrusive game-based assessments sometimes referred to as stealth assessments (V. Shute *et al.*, 2017). It has been used for game-based assessments in subjects like physics (Kim *et al.*, 2016), calculus (Smith *et al.*, 2019) and 21st-century skills (Sweet & Rupp, 2012). The design of assessments for immersive environments using ECD is nontrivial. It is complex and time-consuming and requires expertise in advanced statistics and machine-learning skills (Kim *et al.*, 2016; Westera, 2019; Westera *et al.*, 2020). These are possible reasons for its limited adoption, and the wider preference for traditional assessment types.

6.4.2 Constructive Alignment (CA)

CA is the second principle upon which GBAF was designed. Underpinned by the constructivist learning theory, CA works on the idea that students learn by constructing knowledge through active engagement in the learning environment (Biggs, 2003). The fundamental principle of CA is that intended learning outcomes should be aligned with learning activities and assessment tasks, as shown in figure 6.3.

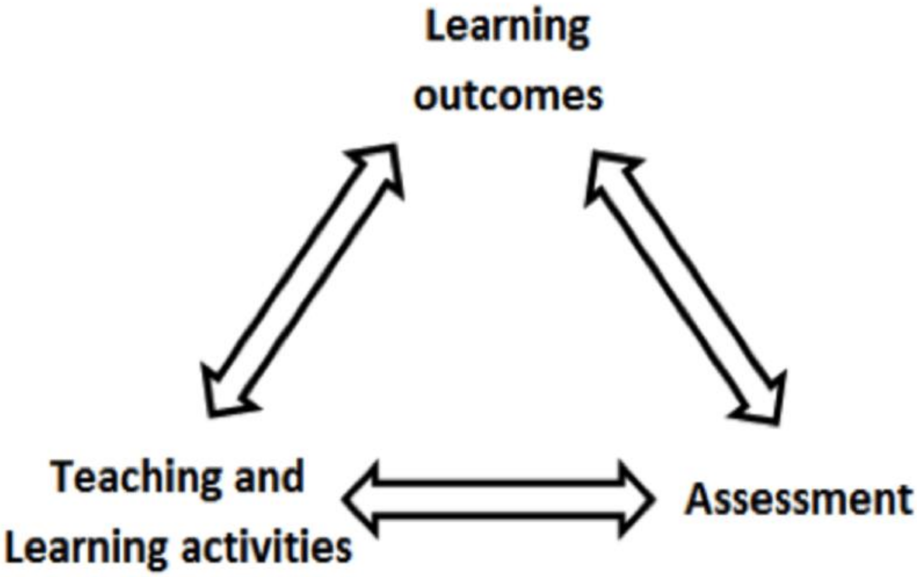


Figure 6.3: Construct Alignment principle (Biggs, 1999).

Biggs proposed four steps to ensure the alignment of all components of the system. First, intended learning outcomes (ILOs) have to be defined following an appropriate taxonomy such as the SOLO taxonomy or Bloom’s taxonomy (Biggs & Tang, 2010). It is also considered important to distinguish between the types of knowledge to be assessed. Two fundamental types of knowledge that need to be distinguished are declarative and functioning knowledge. While the latter is a function of the actions of students, the former is not. Each ILO is to be written with appropriate verbs to indicate standards of achievement (Biggs & Tang, 2010). The second step involves the choice of teaching and learning activities. Although lectures and tutorials, which mostly require passive listening from students, are commonplace in HE, other L&T activities that offer active and engaging environments are recommended. The third step is to design the assessment tasks. Assessment tasks should be aligned with one or more ILOs. These tasks should require students to use the operative verbs in the ILO. An L&T activity that is itself the assessment (such as in games or problem-based learning) offers the best forms of alignment (Biggs & Tang, 2010). Biggs and Tang also argue that assessment tasks are best when they are authentic to the discipline. The last of the four steps in designing assessments using the CA principle is the development of grading criteria.

The CA principle is widely used by educators in HE for curriculum and instructional designs. It is the main principle required for programme specification, assessment criteria and statement of learning outcomes (Ali, 2018). CA provides a logical, effective and familiar principle for assessment design that would allow academic practitioners with or without game-based learning experience to design valid assessments around game tasks. CA complements the ECD framework by emphasising strong connections between ILOs, assessment tasks and L&T activities, factors that are crucial to the design of classroom instructional activities. Whereas CA offers educators familiar guidelines for assessment designs, the ECD has a wider application in the field of games-based assessment.

6.5 Game-Based Assessment Framework (GBAF)

The GBAF draws on the principles of ECD and CA to provide educators, researchers and game designers with an assessment framework that can be applied to immersive learning environments. It also draws on the author’s experience measuring the performance of students in CosmiClean game in Chapters 4 and 5. The challenges encountered in aligning measured learning outcomes with game tasks informed aspects of this framework. As shown in figure 6.4, the GBAF is made up of units outlined in steps.

6.5.1 Step 1: Overall objectives

Immersive learning applications are usually designed to teach a specific topic, subject or concept. To show proficiency level in the subject, students would need to complete sets of tasks in the immersive environment. The actions of students while completing the tasks are used to infer their competencies. Overall objectives describe broadly the purpose of the assessment and in general terms, the competencies assessed. Both the overall learning objectives and the ILOs should also be drafted at this stage. A clear articulation of the objectives of an assessment in an immersive environment is critical to identifying an appropriate immersive application. This can also provide students with information about the learning and performance expectations, in addition to demonstrating the purpose and relevance of the immersive tool.

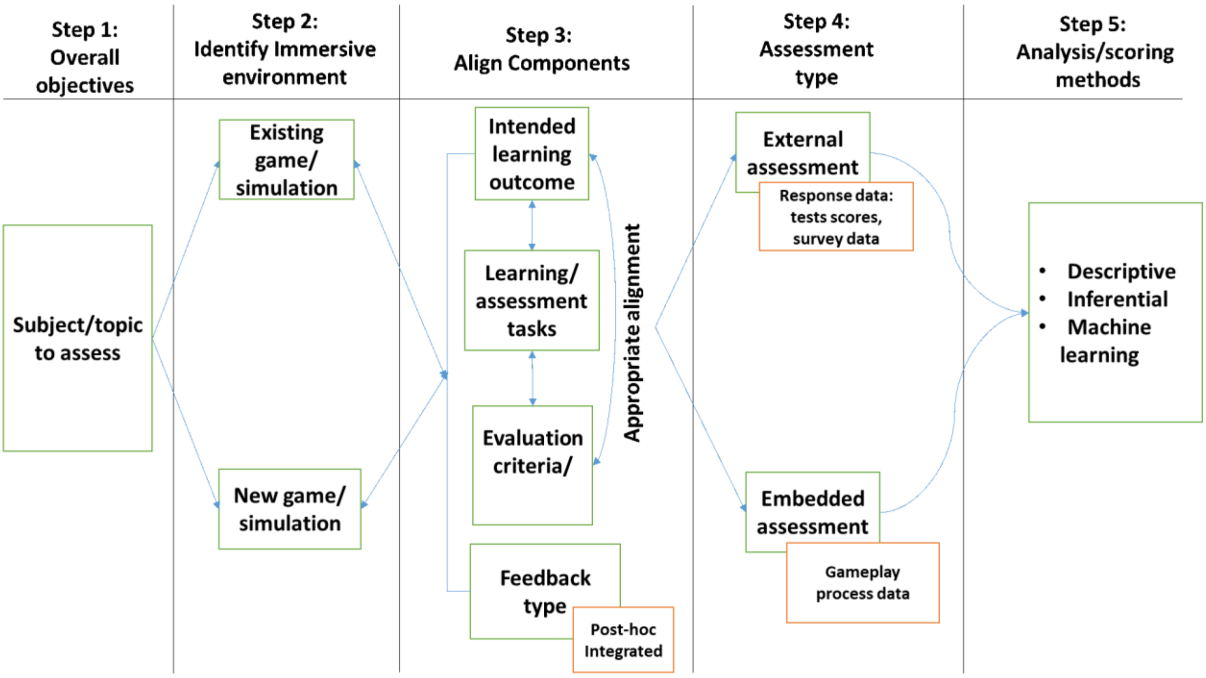


Figure 6.4: Game-based assessment framework (GBAF).

6.5.2 Step 2: Identify an immersive learning environment

In this step, the immersive technology to be used is identified. Whether a DG, VR or AR application, an educator would need to decide on an appropriate environment to use based on the specified objectives. Additionally, educators would have to decide whether to use a pre-existing application or whether it will be more appropriate to create an entirely new immersive learning environment. The decision guidelines presented in figure 6.5 outline decision-making considerations when choosing a pre-existing immersive learning application. For most educators, seeking suitable pre-existing immersive applications for classroom use would be most convenient given the time, expertise and financial costs required to create a new immersive environment. However, finding a pre-existing environment that meets all the needs of the educator might be rather difficult. Appropriate alignment between ILOs, tasks in the game or simulation, and the grading criteria are crucial. When an application fails to account for all desired ILOs but sufficiently covers a good proportion of them, educators might want to consider combining the game-based assessment with other assessment types.

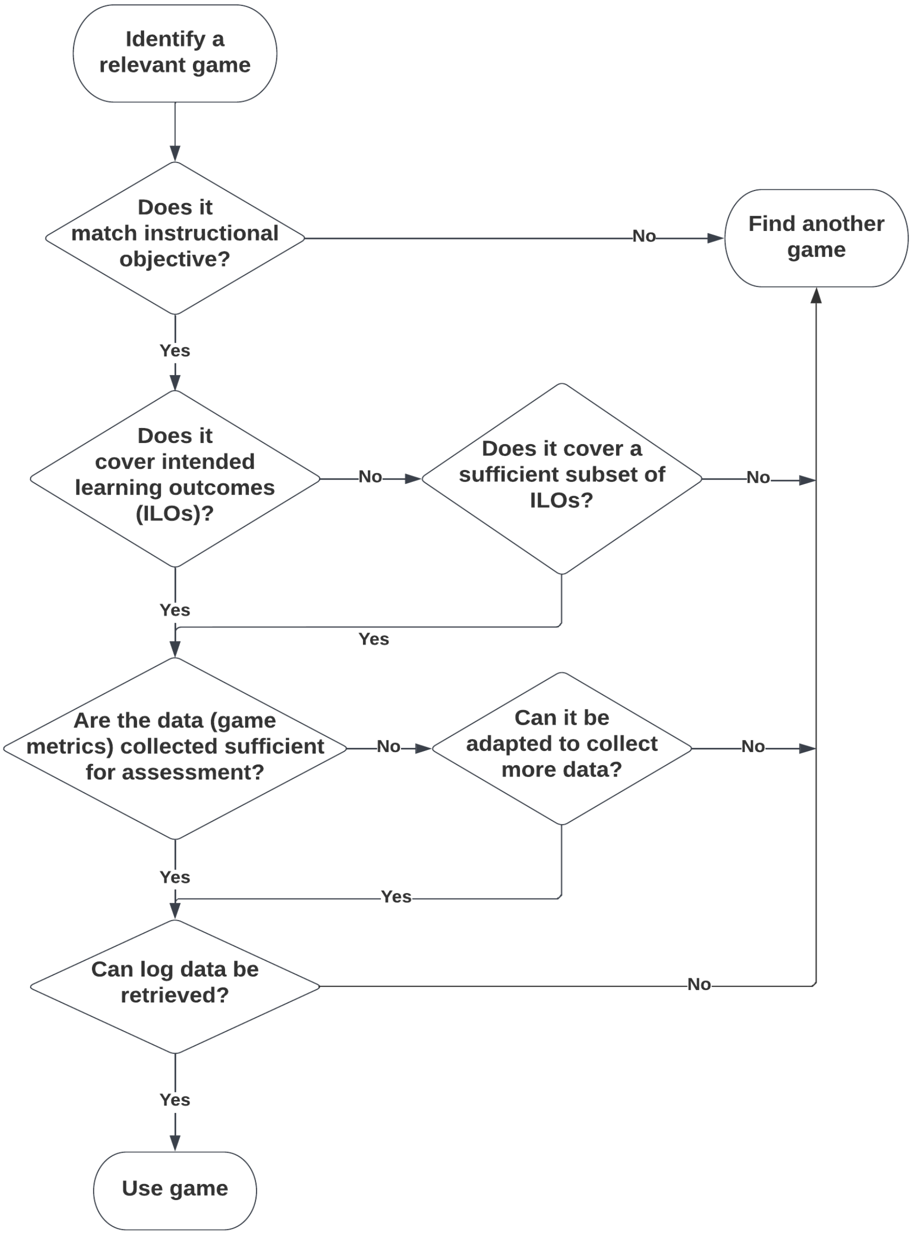


Figure 6.5: Decision-making process for using pre-existing immersive learning applications for assessment.

6.5.3 Step 3: Align components for assessment

Having decided on the most appropriate immersive application to use, the next step is to ensure that all the assessment components fit together. To begin, a list of the game or simulation tasks of the chosen application should be made. Where a new application is to be designed, tasks that would allow students to demonstrate the desired competencies should be designed. Grading criteria for determining the performance of learners based on their actions should also be specified. As demonstrated in figure 6.6, both game tasks and grading criteria components of the assessment must be aligned with the ILOs. Lastly, educators should at this stage think about how feedback would be given to learners: at the end of gameplay (post-hoc) or during gameplay (integrated) or both.

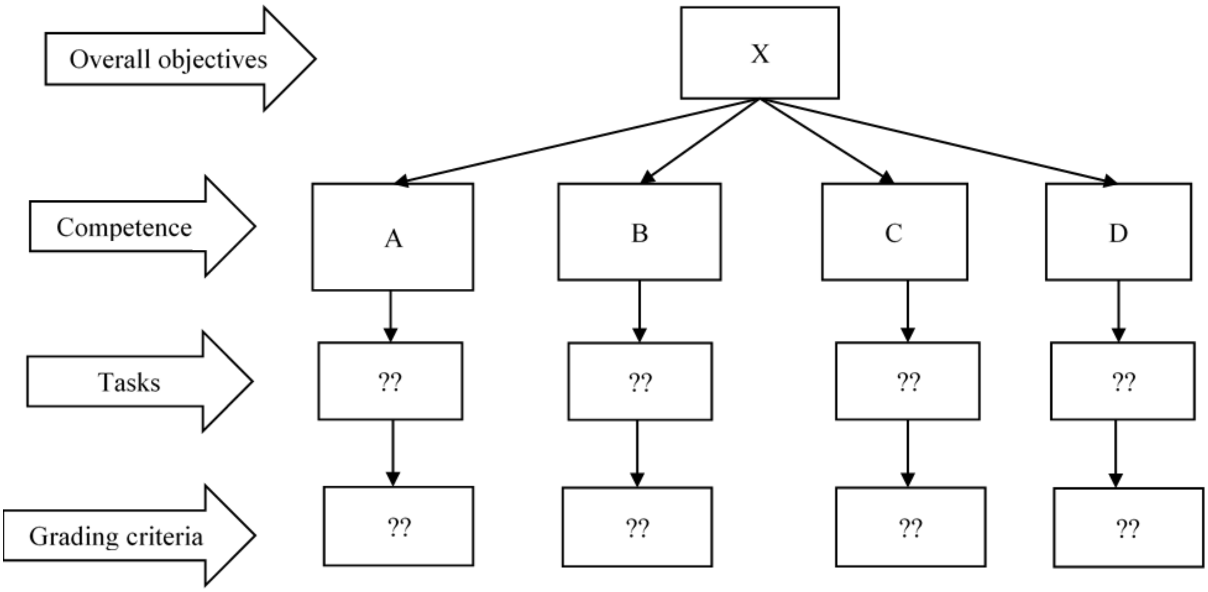


Figure 6.6: Component of the GBAF.

Competence: this describes the specific knowledge, skills, or expertise that students are expected to acquire from a learning activity. For performance-based assessments applicable to immersive environments, statements of ILOs should focus on how students will be able to apply their knowledge in the simulated real-world context. Both the SOLO (Biggs & Collis, 1982) and Bloom's (Anderson, 2013) taxonomies provide frameworks for instructional designs as well as guides for assessment designs. Bloom's taxonomy, however, is well known and used in educational settings as a way of classifying learning goals in terms of the complexity of action or behaviour. It is believed to capture nearly all possible cognitive educational objectives making it very useful for educators *(*Moseley *et al.*, 2005). When designing assessments around immersive learning tasks, the ILOs to be assessed should inform expectations in terms of what students ought *to do* to be considered successful. Each ILO should be written with proper consideration for the available tasks in the adopted immersive learning environment. This is often less complicated when designing a new application because of the flexibility to design tasks around ILOs of interest.

Tasks: immersive technologies offer active learning environments where students interact with game elements or collaboratively with each other to complete given tasks. Unlike conventional assessment tasks that require students to respond to questions, game-based assessment tasks are performance-based. Game tasks constitute activities that require students *to do*, that is, to perform actions in realistic settings. One game task could elicit numerous competencies in students, and thus could be used to measure more than one ILO. Adequate alignment between ILOs and game tasks is crucial to the design of a valid game-based assessment. Available game tasks should sufficiently measure ILOs by requiring students to perform actions that would elicit their level of knowledge of the measured ILOs.

Grading criteria: as with traditional classroom assessments, determining grading methods is necessary if grades are to be awarded. The grading methods are the criteria or formulas for quantifying the competency level of students based on their performance on the tasks. They are used to award marks or grades to students. In game-based assessments where competence is assessed based on actions (or inactions) in complex environments, grading criteria must account for these complexities. Instead of simply grading by correct or incorrect answers chosen or given by students, speed of response, the efficiency of solutions, errors made, hints requested and other variables that facilitate authentic assessments should be considered for grading. Lastly, a scheme outlining the scoring strategy and quality criteria for each performance rating would generally enhance the scoring process. It is good practice to develop grading criteria to ensure reliability and objectivity throughout the grading process (Dawson, 2017; Jonsson & Svingby, 2007).

Feedback type: while alignment of feedback with competence, tasks and grading criteria is not necessary, consideration should be given to the method of providing feedback to students. Feedback can be immediate or delayed (Chaudy & Connolly, 2019), and internal or external to the learning environment. Immersive technologies promote the delivery of immediate personalised feedback to learners. However, when using a pre-existing application, efforts should be made to understand the relevance of the available integrated feedback components. When designing a new application, necessary feedback mechanisms should be built into the environment to provide timely feedback to students. Where feedback cannot be built into the immersive learning environment of choice or when it is considered inappropriate to provide feedback during gameplay as it might interfere with assessments, delayed feedback can be provided either in the immersive environment or in person. Post-gameplay (post-hoc) debriefing sessions that allow for reflection on the game tasks and play actions taken offer good opportunities for educators to address misconceptions and correct errors (Crookall, 2010).

6.5.4 Step 4: Assessment type

In Chapter 2 of this thesis, the different methods of assessments relevant to immersive learning technologies were discussed. To measure the learning performance of students when using immersive technologies, it is recommended to opt for embedded assessment methods as these offer benefits difficult to achieve with external assessments (Kim *et al.*, 2016; Shute *et al.*, 2017). In cases where educators cannot design embedded assessments, external assessment methods may be considered. However, the assessment items or questions should be designed to be authentic and contextual. This could mean modelling assessment items around gameplay tasks. The application of GBAF to an external assessment design is presented in Chapter 7.

6.5.5 Step 5: Analysis and scoring methods

Lastly, educators need to determine the statistical methods that will be used to analyse the performance of students. This could range from descriptive statistics such as sums, percentages and means, to inferential statistics and machine learning techniques. For performance scoring purposes, descriptive statistics should be considered given that they are easy to compute and are also frequently used by educators to score traditional assessments. Inferential statistics and machine learning techniques are often used for research purposes to compare groups and predict future events. Although more complicated to carry out compared to descriptive statistics, these statistical methods can provide additional pieces of information. Larger sample sizes (in 100s) are usually required for these analyses. The recommendation to educators would be to use simple but appropriate statistical methods that they are comfortable with. When these are insufficient for the desired assessment, other sophisticated statistical methods should be considered. When integrating assessment into the design of an immersive learning application, statistical equations can be incorporated, such that the data analyses are automatically carried out by the computer, and scores awarded to students immediately after gameplay. Technical details of how this can be implemented are beyond the scope of this research.

6.6 Applying GBAF to the Design of Assessments for a DG

In this section, the design of assessment for an educational DG using the GBAF is presented. CosmiClean, the same game used in previous chapters (see Chapter 4 for more details) is adopted here to show how the GBAF can be applied to an assessment design. Data from the gameplay of students presented in Chapters 4 and 5 were used for this analysis. Data were anonymised and ethical approval was granted for the use of the data for this purpose. First, the decision-making guideline for choosing games for assessment, shown in figure 6.5, was used to establish the suitability of CosmiClean for the given assessment purpose. The overall objective of the assessment was to measure the proficiency of chemical engineering students in the separation operations of solid materials. A DG that is able to assess proficiency in separating materials of different characteristics using different equipment was considered ideal. To enable the identification of a suitable game, the following broadly stated learning outcomes were set:

1. Students should be able to separate materials by their physical properties using different separation equipment.
2. Students should be able to separate materials by their chemical properties using different separation equipment.

To determine the suitability of CosmiClean for these purposes, the author first played the game to evaluate its appropriateness for the specified ILOs. On completing the game tasks, it was found that CosmiClean provided opportunities to assess proficiency levels in the separation of materials by their physical but not their chemical properties. A broad range of physical properties is covered by the game as well as several types of separation equipment. Although only some of the learning outcomes of interest are covered by the game, CosmiClean was considered useful for assessment at that point.

The GBAF was applied to the design of embedded assessments for CosmiClean. The embedded assessment utilised log data for scoring the performance of students. Adopting the GBAF, all five steps were considered in the design of assessments for the first 25 levels of CosmiClean. Steps 1 and 2 were carried out during the evaluation of the appropriateness of CosmiClean for separation operations. As earlier mentioned, steps 4 and 5 are a function of the type of immersive environment selected for assessment. For the present case study, embedded assessments are carried out using the gameplay log data of students. Descriptive statistics of their performance in the game will be presented. For step 3, it is necessary to elaborate on the process followed to ensure alignment between the ILOs, tasks and grading criteria. The alignment of assessment components in accordance with the GBAF is shown in figure 6.7 with the overall objective of the assessment and the ILOs aligned with the game tasks and proposed scoring methods.

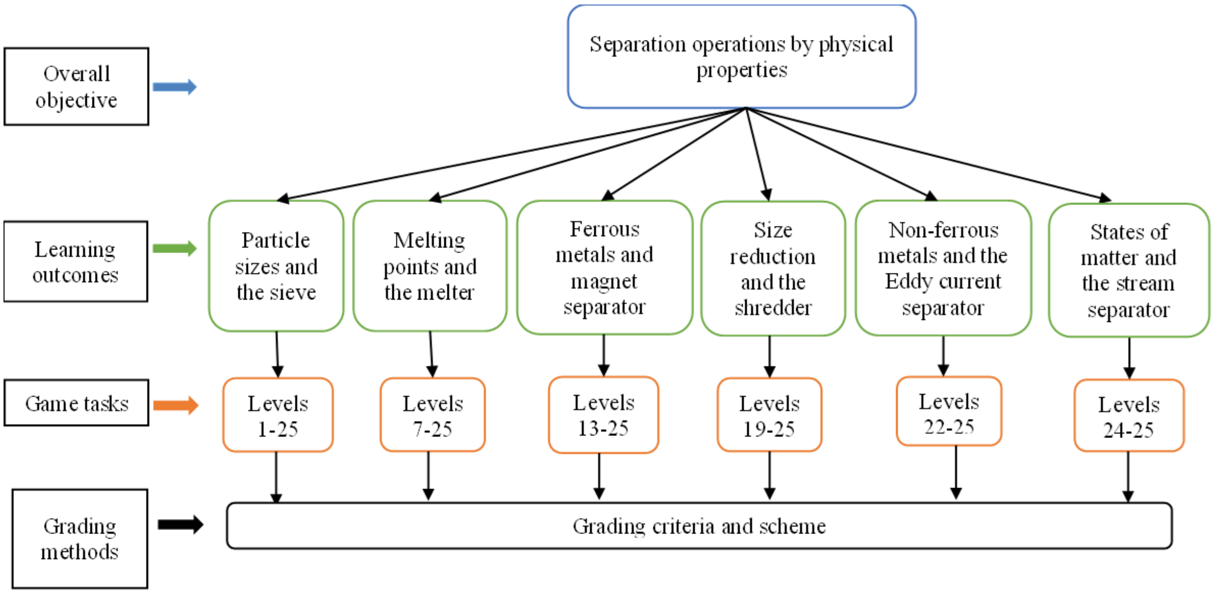


Figure 6.7: Assessment components alignment for CosmiClean.

Learning outcomes: the following ILOs were assessed and these all measure "applying" as well as conceptual and procedural knowledge dimensions of Bloom's taxonomy (Anderson, 2013). During the gameplay, students should be able to

1. Correctly configure the sieve to separate particles of different sizes.
2. Correctly configure the melter to separate particles based on their melting points.
3. Correctly configure the magnetic separator to separate ferrous metals from non-metals.
4. Correctly configure the shredder to separate heterogeneous materials by size reduction.
5. Correctly configure the Eddy Current to separate non-ferrous metals from ferrous metals.
6. Correctly use the stream separator to separate mixtures of liquid and solid particles.

Game tasks**:** the tasks for each level of the game are similar but with increasing complexity. Students were required to drag and drop processors, conveyors and receptors that collect separated particles in the appropriate positions. As depicted in figure 6.7, for levels 1 to 6 of the game, students were assessed on their knowledge of separation by size. At these levels of the game, the sieve was introduced and students were expected to place the required number of sieves and receptors at appropriate positions, and then correctly configure them, that is, set the correct mesh sizes to separate the waste particles presented. At levels 7 to 12, students were assessed on their ability to correctly separate particles by their melting temperatures. This involves correctly configuring the melters, setting these to the appropriate temperatures, to separate particles with different melting temperatures. In addition to using the melters, students were required to incorporate the sieves where necessary to achieve complete separation of the presented materials. The knowledge of the separation of metals from non-metals was assessed in levels 13 to 18, along with the ability to combine these with previously introduced separators to achieve sustainable separation. It must be noted that the energy consumption rates of each processor differ and were presented to students during the gameplay. With this information, students could apply more sustainable solutions when completing the tasks. At levels 19-21, separation by size reduction was assessed along with the ability to combine the sieves, melters and/or magnets to achieve complete separation of the presented waste materials. Similarly, at levels 22-23 and 24-25, students were assessed on their ability to separate particles by their ferrous and non-ferrous metal characteristics, respectively. They were also assessed on their ability to incorporate any of the previously introduced processors to sustainably achieve separation.

Grading criteria: given the many possible solutions to the tasks in CosmiClean, the author had to note the most optimal solution to each task. Also, to identify the different possible solutions to the tasks, data from previous gameplays of 146 university students were assessed. These were compared with the solutions of the author to identify the “optimal” performance for each level of the game. Using the game metrics shown in table 6.1, the 25th 50th and 90th percentile values of the data were calculated in Microsoft Excel. The solutions of students in the 90th percentile were most similar to those of the author, hence were labelled the “Optimal Performance” group. The performance of students in the 50th and 25th percentile were labelled “Competent Performance” and “Novice Performance”, respectively. An assessment of the data showed that, whereas optimal performers (students in the Optimal Performance group) used the least number of resources on each task, novice performers (students in the Novice Performance group) consumed the most resources on each task. The performance of the competent performers (students in the Competent Performance group) was considered average as shown in table 6.1. The values presented are the upper limits of the range of scores for each metric, which means that, for level 1, time on task for optimal performers is ≤26 seconds, for competent performers >26 seconds but ≤40 seconds, and for novice performers >40 seconds but ≤90 seconds, respectively. The metrics chosen here for this analysis are solely for illustrative and research purposes. The choice of metrics could vary based on several factors such as the competencies measured, the student characteristics and the metrics captured by the game.

Table 6.1: Grading scheme for embedded assessment of performance in CosmiClean.

Game levels	**Scoring metrics**	**Optimal Performance (Top 25%)**	**Competent Performance (Top 50%)**	**Novice Performance (Top 90%)**
1	Time on task[s]	26	40	90
	Energy consumed	0	0	0
	No. of Equipment units	4	4	4
2	Time on task[s]	30	45	100
	Energy consumed	24	24	24
	No. of Equipment units	4	4	4
3	Time on task[s]	65	90	230
	Energy consumed	24	24	24
	No. of Equipment units	3	4	5
4	Time on task[s]	70	125	450
	Energy consumed	16	16	16
	No. of Equipment units	3	4	5
5	Time on task[s]	80	110	330
	Energy consumed	44	44	48
	No. of Equipment units	5	6	7
6	Time on task[s]	110	165	400
	Energy consumed	36	36	36
	No. of Equipment units	7	9	11
7	Time on task[s]	80	100	300
	Energy consumed	32	32	40
	No. of Equipment units	3	3	4
8	Time on task[s]	50	70	160
	Energy consumed	22	22	30
	No. of Equipment units	3	3	3
9	Time on task[s]	70	90	170
	Energy consumed	44	44	50
	No. of Equipment units	5	5	6
10	Time on task[s]	80	110	250
	Energy consumed	80	80	85

	No. of Equipment units	5	5	6
11	Time on task[s]	80	120	240
	Energy consumed	54	54	60
	No. of Equipment units	5	5	7
12	Time on task[s]	60	70	180
	Energy consumed	81	81	95
	No. of Equipment units	5	5	6
13	Time on task[s]	40	50	100
	Energy consumed	60	60	60
	No. of Equipment units	3	3	3
14	Time on task[s]	30	40	70
	Energy consumed	60	60	60
	No. of Equipment units	3	3	3
15	Time on task[s]	50	70	150
	Energy consumed	25	42	53
	No. of Equipment units	5	5	5
16	Time on task[s]	80	140	350
	Energy consumed	45	46	50
	No. of Equipment units	5	5	6
17	Time on task[s]	65	100	340
	Energy consumed	64	66	74
	No. of Equipment units	5	5	6
18	Time on task[s]	110	140	300
	Energy consumed	77	88	120
	No. of Equipment units	7	10	12
19	Time on task[s]	60	70	170
	Energy consumed	20	160	210
	No. of Equipment units	4	4	5
20	Time on task[s]	60	85	170
	Energy consumed	20	104	1120
	No. of Equipment units	5	6	6
21	Time on task[s]	65	80	150
	Energy consumed	24	24	105
	No. of Equipment units	5	5	6
22	Time on task[s]	40	55	90
	Energy consumed	4	60	60

	No. of Equipment units	3	3	3
23	Time on task[s]	60	80	170
	Energy consumed	9	135	135
	No. of Equipment units	5	5	5
24	Time on task[s]	50	60	110
	Energy consumed	40	40	40
	No. of Equipment units	3	3	3
25	Time on task[s]	40	60	120
	Energy consumed	32	38	60
	No. of Equipment units	5	5	5

6.7 Evaluating the GBAF

The assessment framework provided valuable guidance that enabled the design of assessments for the desired ILOs as well as the identification of a suitable game for the intended purpose. It also allowed the linking of ILOs to game tasks, a very important factor when designing assessments for immersive learning environments. It provided a structured and easy approach to the design of assessments for a DG. Unlike the ECD framework, the GBAF does not require the use of advanced statistical skills to determine performance in the DG environment. The GBAF provides an intuitive approach to aligning ILOs with game tasks. Its application to CosmiClean assessment design demonstrates that the GBAF can be a practical framework for the design of assessments for immersive learning environments. It can serve as a structured approach for designing authentic assessments of learning based on the actions of students in immersive environments. It has been found that the performance of students in game-based assessments is positively correlated with their performance in multiple-choice tests (Sanchez *et al.*, 2022).

Although the GBAF provided a structured approach to assessment design around a DG's tasks, it was observed that GBAF did not consider the scoring metrics needed to ascertain the performance of students. During the design of the CosmiClean assessment, it was unclear what to consider for the measurement of the performance of students. The retrieved log files contained several data on the gameplay of students. These include duration of gameplay, dates, times, number of repetitions made, internet protocol (IP) addresses of players, *etc*. The decision on which sets of metrics to include and whether to include all these metrics in the assessment was not accounted for. The GBAF offered no guidance on this, which made its use for assessment design less than efficient. Integrating considerations for scoring metrics into the framework, and providing practical guidance on how to identify potentially relevant scoring metrics from the game metrics, would certainly improve its usefulness. Game metrics differ from one immersive environment to another, hence an understanding of how to build in or select metrics that appropriately align with the other assessment components is crucial.

In addition to the observed limitation of GBAF made by the author, an evaluation was carried out with a member of staff who was interested in using immersive learning technologies for chemistry laboratory practices. At the time of this study, it was not possible to recruit more educators for this evaluation because staff were on holiday for the summer. Only one educator responded to the invitation and arranged to meet for the session. This educator was chosen for this evaluation because of their interest in the applications of immersive technologies in education. In an informal interview with them, the assessment framework was presented. They were first presented with the framework and then asked whether they understood it. They were also asked if they could design a hypothetical game-based assessment using the framework. On scanning through the sheet of paper with the framework, their first response was, "*oh gosh! It looks quite complicated. Could you explain it to me*?" This response was not the expected response for a user-friendly assessment framework. It immediately highlighted another limitation of the GBAF. However, after explaining how the framework works and how it can be applied to assessment designs for immersive learning applications, the educator was able to describe how they would potentially apply the GBAF to the design of assessments, pointing at the different steps of the framework. At the end of the discussion, they recommended simplifying the framework to make adoption easier.

As it was difficult to schedule more evaluation opportunities with educators at this time, the supervisors of the author of this thesis provided some feedback. Their feedback was similar to that of the independent educator, recommending a simplification of the framework. They both found it quite complicated to digest as it had a lot of information presented within it. Based on this feedback and the observation made during the design of the embedded assessment for CosmiClean, a revision of the GBAF was deemed necessary. The revision of the assessment framework is presented in Chapter 7.

6.8 Limitations

The major limitation of this chapter is the number of educators who provided feedback on the GBAF. An evaluation study involving 5-7 educators would have provided better feedback and insights into the limitations and potentials of the GBAF.

6.9 Conclusions

This chapter outlines the development, application and evaluation of the GBAF. It also provides decision-making process guidelines that can help educators with the choice of immersive learning environments to use for a given assessment purpose. The GBAF, founded on the ECD and CA principles, was designed to provide a structured and logical basis for the design of valid assessments for immersive learning. Its application to the design of assessments for CosmiClean showed that it does enable the design of assessments around DG tasks. Although it was relatively useful for an assessment design, it missed out on a crucial component required for the design of assessments for immersive learning. While the outcome of the evaluation of the framework is limited by the small sample size of evaluators, the findings nonetheless indicated the need for a simplification of the framework. The GBAF is intended to be easy to use and effective for assessment designs, hence, the feedback concerning its complexity is considered significant. In the next chapter, a revised GBAF will be presented.

6.9 Note: Part of this chapter has been submitted to **British Journal of Educational Technology -** Udeozor, C., Abegão, F. R., & Glassey, J. (Submitted). Measuring Learning in Digital Games: Applying the Game-Based Assessment Framework. *British Journal of Educational Technology***.**

Chapter 7. Revised Game-Based Assessment Framework: Evaluation and Applications

This chapter presents the revision of the game-based assessment framework (GBAF) following the concerns raised by educators and the observations made by the author. It also presents an evaluation of the revised framework (GBAF 2) by three educators. The GBAF described in Chapter 6 outlined the steps necessary to carry out assessments for immersive immersive learning. One of those steps considered crucial for the design of assessments around tasks in immersive learning environments was the 3rd step which emphasised the alignment of assessment components. The current chapter presents a decoupling of the GBAF into two parts: the GBAF 2 and the assessment design steps. The assessment design steps outline the approach to consider when conducting assessments with immersive learning applications. The GBAF 2 on the other hand, highlights components or factors that must be taken into account when designing assessments to measure learning in immersive environments. It also demonstrates the relationship and alignments needed between components of assessments for an effective assessment design. The application of the GBAF 2 to the design of assessments for a virtual reality (VR) game, an augmented reality (AR) game and an educational digital game (DG) are also presented. The robustness of the framework for external and embedded assessment designs is demonstrated by its application to the design of both embedded assessments for the CosmiClean game in Chapter 6 and the design of an external assessment in the current chapter. An evaluation of the framework by three educators from the school of engineering showed that GBAF 2 is easy to use. Its successful application to the design of assessments for three different environments also shows that it is more efficient for assessment designs compared to the GBAF. The GBAF 2 has the potential to provide a structured and unified framework that would enable the design of assessments when using immersive learning technologies.

7.1 Introduction

The GBAF was developed to respond to the lack of a structured approach for designing assessments for immersive learning environments. The growing adoption of immersive learning technologies in higher education (HE) requires corresponding adjustments to be made to current educational practices. The assessment of learning is a vital part of learning and teaching (L&T) in every formal educational setting. The critical role of assessments in the learning and future goals of HE students means that high standards of assessment practices must be adhered to (Lock *et al.*, 2018). Although established guidelines such as those provided by the Constructive Alignment (CA) principles (Biggs & Tang, 2010), the SOLO taxonomy (Biggs & Collis, 1982), and Bloom's taxonomy (Anderson, 2013) are useful for traditional assessment practices, their effectiveness for non-traditional assessment practices is yet to be determined. Efforts have been made by Loh (2012) and Mislevy, Steinberg and Almond (2003) to provide assessment frameworks for immersive learning environments. Loh's framework emphasises process data collection and analyses with little to no guidance on the design and implementation of assessments in these environments. Mislevy, Steinberg and Almond (2003) on the other hand provided a framework that is useful for the design, implementation and scoring of assessments in complex learning environments such as immersive environments. However, the application of this framework to classroom assessment practices has been limited, due to the complexity in design and the advanced statistical skills required for it (Westera, 2019; Westera *et al.*, 2020; Kim *et al.*, 2016). Additionally, the application of the framework has been restricted to the design of embedded assessments (see Kerr and Chung, 2012; Almond *et al.*, 2014; Jaffal and Wloka, 2015; Shute, Rahimi and Emihovich, 2017) making it less useful to educators who are less likely to have the time and resources required to design new immersive applications for classroom pedagogy. To bridge this gap in the literature, the GBAF was developed to help educators with the design and implementation of assessments when using immersive learning applications.

The GBAF developed in Chapter 6 builds upon two established conceptual frameworks – the CA principle (Biggs, 2003) and the Evidence Centred Design (ECD) framework (Mislevy, Steinberg, *et al.*, 2003). It is made up of five units that outline the necessary steps needed to create assessments for immersive learning environments. The 3rd step, shown in figure 6.4, is considered the most important as it emphasises the relationship and alignment between the components necessary for the design of assessments for measuring learning in immersive environments. While the GBAF enabled the design of assessments around the tasks of an educational DG, CosmiClean, it was found to miss an important factor, which is, the scoring metrics, necessary for ascertaining performance in immersive environments. Additionally, feedback from educators who evaluated the GBAF showed that the framework was not easy to understand. This feedback and the observations made by the author warranted a revision of the framework to address these issues. Therefore, the goal of this chapter is to present a revised version of the GBAF that addresses all the necessary components for assessments relevant to immersive environments, and one that is easier to understand and use.

7.2 Revision of the GBAF

As earlier stated, the most important part of the GBAF is the alignment of assessment components addressed in step 3 of the framework. This step is considered important because it highlights the relationships between assessment components and provides guidance for establishing alignment between those. Although the other steps are relevant for assessments in immersive learning environments, they mainly describe the procedures needed to implement assessments when teaching with immersive learning applications. An assessment framework for immersive learning environments should be able to describe the connections between important factors necessary for the design of assessment tasks. These links between components necessary for designing assessment tasks are often the basis of other assessment frameworks such as the ECD framework. To emphasise these important components and their relationships, the 3rd step of the GBAF is extracted, modified and elaborated on as the revised game-based assessment framework (GBAF 2). This thus reduces the original GBAF to assessment design and implementation steps, and the GBAF 2.

7.2.1 Assessment design and implementation steps

The assessment design steps shown in figure 7.1 are made up of four iterative steps that demonstrate the process required to identify, design and choose assessment methods for immersive learning environments.

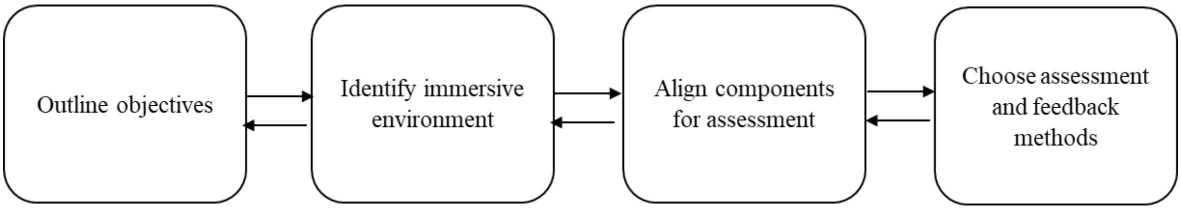


Figure 7.1: Steps for assessment design for immersive learning.

Outline objectives: as detailed in section 6.5 of Chapter 6, the first step to creating assessments for these environments is to define the overall objectives of the assessment. In this step, a clear articulation of the objectives of the intended assessment is first specified to enable the design or identification of relevant immersive learning environments. The statements of objectives of assessments can also provide necessary information to students about the learning expectations and the relevance of the immersive learning technology. Additionally, a draft of the intended learning outcomes (ILOs) is recommended at this stage as it would enhance the process of selecting an appropriate application for the assessment.

Identify immersive environment: the identification of an appropriate immersive learning environment comes next. This can be in the form of a DG, VR or AR application that meets the intended purpose detailed in the statement of the overall objectives. Figure 6.5 of Chapter 6 provides decision-making guidelines that can be used by educators to determine the appropriateness of a pre-existing immersive application for an assessment purpose. Where an educator fails to identify an appropriate pre-existing application, the design of an entirely new environment should be considered.

Align components for assessment: once an environment is identified, connections should be made between the overall objectives and ILOs initially defined, with the tasks in the immersive environment, the scoring metrics and the grading method. This step will be addressed in more detail in subsection 7.2.2.

Choose assessment and feedback methods: the two major assessment methods applicable to immersive learning environments are external and embedded assessments described in Chapter 2. The use of embedded assessment methods is recommended because they provide performance-based information on the actions of students that are considered authentic assessments. Nonetheless, external assessments are still widely preferred. The comparative ease of design of external assessments, and the challenges of finding an application that collects and stores all relevant game metrics needed to measure specific ILOs, mean that educators may have to rely on external assessment methods. When properly designed, external assessments can be performance-based and authentic. The application of GBAF 2 to the design of an external assessment for a DG is presented in subsection 7.4.3. In addition to determining the appropriate assessment method to use, the type of feedback to be provided to students should be considered. When a new immersive environment is designed for learning and assessment, immediate feedback and adaptive interactions should be incorporated as scaffolding measures to reduce the extraneous cognitive demands on students and inherently enhance learning (Desurvire *et al.*, 2004; Kalyuga & Plass, 2009). Delayed feedback could also be provided to students either individually or in groups during debriefing sessions. This is particularly necessary when pre-existing applications that might not provide all the necessary feedback to students are used.

7.2.2 The revised game-based assessment framework (GBAF 2)

The steps outlined in subsection 7.2.1 above demonstrate the entire process of creating an assessment for immersive learning. The GBAF 2, on the other hand, focuses on demonstrating the important components necessary for these assessments, as well as the connections between them. The GBAF 2 comprises the overall objective of an assessment, the ILOs, the tasks in the immersive environment of choice, the scoring metrics and the grading methods. As shown in figure 7.2, the emphasis is on the alignment of all five components.

Overall objective: the overall objective component of the GBAF 2 describes the goal of the assessment in terms of the module, subject or topic of interest. Detailing the objectives of the assessments in relation to the curriculum is the first step toward the design and alignment of all the components needed for assessments.

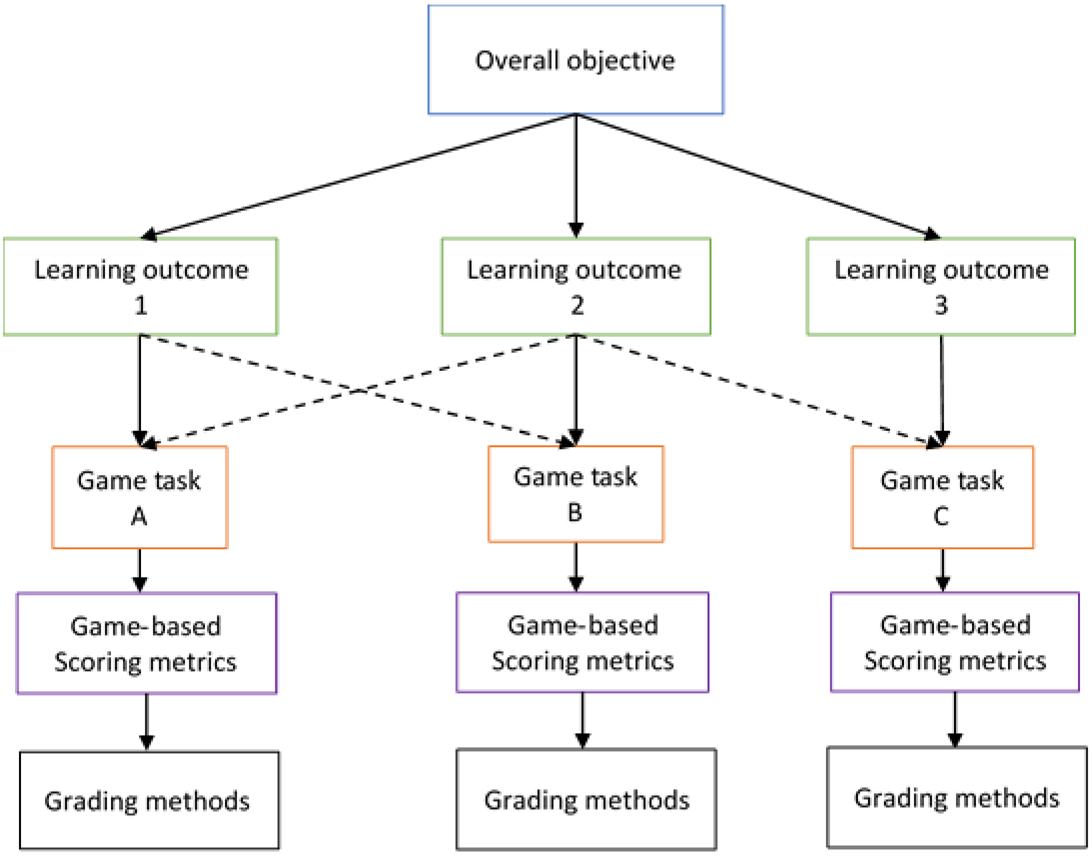


Figure 7.2: The revised game-based assessment framework (GBAF 2).

Learning outcomes: this describes the specific knowledge, skills, or expertise that students are expected to acquire from an immersive learning activity. These ILOs are expressed as statements of what students should be able to do in the immersive environment and are often drafted using Bloom's taxonomy (Anderson, 2013). When designing assessments around immersive learning tasks, the ILOs to be assessed should inform expectations in terms of what students ought *to do* to be considered successful. Each ILO should be written with proper consideration for the available tasks in the adopted immersive learning application. This is often less complicated when designing a new immersive environment because of the flexibility to design game tasks around ILOs of interest.

Tasks: tasks in immersive learning environments constitute activities that require students to perform actions or solve real-life problems in realistic settings. One game task could elicit numerous competencies in students and hence could be used to measure more than one ILO. Adequate alignment between ILOs and game tasks is crucial to the design of a valid game-based assessment. Available game tasks should sufficiently measure ILOs by requiring students to perform actions that would elicit their level of knowledge of the measured ILOs.

Scoring metrics: this is the only component absent in the original GBAF. Scoring metrics refer to those game metrics that can be used to assess the performance of students on predefined ILOs. VR, AR and DGs collect and store process or telemetry data of the actions of students that can be used for assessment purposes. This means that real-time process data of students in immersive learning environments are used to infer competencies. These data could be as general as "tasks completed", "time spent on task", "levels completed", or "correct and incorrect answers", or as detailed or grain-sized as "numbers of retries", "materials and hints accessed" or "locations visited".

During the development of a new immersive application, the scoring metrics can be determined and integrated into the game mechanics, however, this flexibility is significantly limited when a pre-existing immersive application is adopted. To ensure that the ILOs of interest can be adequately and sufficiently measured in a pre-existing immersive environment, it is recommended to first outline game metrics associated with each game task. Next is the identification of relevant metrics that would be useful for measuring each ILO (which will serve as the scoring metrics), and the considerations for how these can be used to infer levels of achievements on each ILO. It is worth noting that the quantity and quality of data collected in immersive environments differ. This can pose a major challenge for educators as fewer data might limit the number of ILOs assessed while too much unnecessary data might complicate assessment designs and increase data processing times (Loh, 2009). This challenge is common with pre-existing applications that are designed with no assessment considerations from the outset. Meanwhile, when assessment is embedded during the design phase of the game, data collection can be limited to information pertinent to the assessed ILOs. Nonetheless, identifying relevant metrics in a pre-existing immersive learning environment of choice could require reverse engineering, that is working backwards starting with identifying available game metrics to determining the ILOs to measure.

Grading methods: grading methods are the statistical formulas used to compute the competency levels of the students assessed. When assessing students, actions (or inactions) in the immersive environment are used to infer their competencies, making assessments in these complex environments quite different from traditional exams. An understanding of these differences and the identification of adequate scoring metrics that align with ILOs would inform the choice of relevant grading methods. Descriptive statistics commonly used for grading traditional classroom assessments can be applied to the grading of VR, AR and DG-based assessments. It is considered good practice to develop grading schemes during assessment design because it enhances the scoring process and ensures the reliability and objectivity of the grading process (Dawson, 2017; Jonsson & Svingby, 2007). Grading schemes outline the quality criteria that would determine what grade or performance rating is awarded to students given their performance in the immersive environment.

7.3 Evaluation of GBAF 2

To determine the usability of the GBAF 2 for assessment designs, a qualitative evaluation of the framework was conducted.

7.3.1 Participants

Three educators from the school of engineering took part in the evaluation. They were recruited from the Teaching Action Group at Newcastle University during an Assessment and Feedback session. This group of educators were purposefully recruited for this evaluation because of their interest in assessments and because they were interested in innovative learning approaches. All three educators worked at Newcastle University, UK and have engineering backgrounds with two of these educators having teaching functions at the time of this study. The participants were lecturers in geotechnical engineering, power systems engineering and one L&T engineering education manager.

7.3.2 Data collection and analysis

Semi-structured interviews were conducted in person with each participant. Each session lasted about 25 minutes. Participants responded to a total of eight open-and closed-ended questions outlined below. The first three questions were used to assess their views on teaching and assessment with technologies. The last five questions measured their views on teaching and assessments with immersive learning technologies and their thoughts on the GBAF 2. Responses to the interview questions were manually noted on a Microsoft Word document. Ethical approval was granted for this study. The questions covered were as follows:

1. Do you currently use technologies in your teaching?
2. How about for assessment?
3. Are there any barriers to the use of technologies for teaching and assessment?
4. Have you considered or do you use any of the emerging immersive technologies, *e.g.* VR, AR, and DGs?
5. What would make you use those?
6. How about for assessments?
7. How about this framework? Without being described, do you understand the framework?
8. Can you describe how you can apply it to a hypothetical immersive learning environment?

Data collected were manually analysed on Microsoft Word. First, responses to each question were pulled together and reviewed. A descriptive analysis of these responses was carried out to identify key phrases or sentences that provide valuable information to the corresponding questions. These key points were then summarised to provide answers to each of the questions.

7.3.3 Results

For the first six questions that are not directly linked to the GBAF 2, the points made by the participants and themes identified from their responses are outlined in table 7.1. All the participants reported using technology for teaching. The technologies mentioned include videos, learning management systems, interactive apps and open-source programming environments. They also reported using similar technologies for assessments. The barriers to the use of technologies for teaching and assessments identified in their responses can be grouped as differences in the preferences of students, a lack of necessary technological resources, the usability of the provided technological applications, and accessibility of relevant technologies. One participant stressed the need for unified but relevant technologies across institutions. They emphasised that educators often have to adapt each time they moved from one institution to another and that in most cases, the needed technologies are not always available. They recommended the use of open-source technologies that are easily accessible across institutions and geographical locations. This participant discussed the need for more efficient educator- and student-friendly technologies that would enhance L&T.

Table 7.1: Key points and summary of findings from interviews with educators to evaluate the GBAF 2.

Interview Questions	**Points**	**Summary**
Do you currently use technologies in your teaching?	Yes; canvas, videos, board games. Yes; Blackboard, Slido, Canvas, google Colab. Yes.	N/A
Do you use technologies for assessments?	Yes; quizzes and mini-projects. Yes, Easy for assessment [implementation], collaboration and feedback. Definitely; enhances formative assessments and feedback.	N/A
Are there any barriers to the use of technologies for teaching and assessments?	Not [suitable] for all students. Variety of assessment types necessary. Lack of relevant technological resources and applications; differ from institution to institution. More open-source applications needed. One-size-fits all applications that are not useful. Usability and accessibility.	Students’ preferences Lack of resources Technology usability accessibility

Have you considered or do you use any of the emerging immersive techs, e.g. VR, AR, DGs?	No. Designing these is time-consuming but I would adopt one if I find. No. But should be engaging and interesting to students. Not personally as I do not teach at the moment, but I will recommend it to teaching staff	Positive about their potential
What would make you use immersive technologies?	If I find a suitable one for my module and student cohort, of course I will use it. It must be useful for my module and interesting to the students. Relevant. Accessible.	Relevance to module Accessible
Would you use immersive technologies for assessments?	Yes. If the environment allows it. Possibly. Can be used for assessment if properly developed I would want to explore it.	Openness to explore

-N/A: not applicable

For questions four to eight, the discussions were focused on immersive learning technologies. Although one participant reportedly plays games in VR, none of the participants had used immersive technologies for teaching or assessments. They all, however, had positive impressions of the potential of immersive technologies to improve the learning experiences of students. In all three interviews, there were major discussions about the challenges of adopting immersive technologies for education. The participants all highlighted the time and skills requirements for designing relevant applications for classroom use. They considered these issues the major barrier to using these technologies for teaching. When asked about the use of immersive technologies for assessment, the participants did not show the same level of enthusiasm as they did with immersive technologies for teaching. They seemed open to the idea but also showed some reservations. This was not unexpected, particularly because none of these educators had used these technologies in their teachings. Conceptualising the practicalities of assessing learning in immersive environments would be challenging for anyone with similar limited experiences.

All participants agreed that a framework or guidelines useful for assessments in immersive environments would certainly enhance their adoption. When presented with a diagram of the GBAF 2, each participant took about 40 seconds to look through it while describing how they think it works. One participant thought it was intuitive and easy to understand. Another agreed that all the components were indeed necessary for assessment but described the development of the immersive learning environment as the bottleneck. This participant showed much more interest in the design of the learning environments than in their use for assessment:

> "*… I think this is easy to understand and implement for assessment design. This [pointing at the 'task' component of the GBAF 2] is the biggest problem…How do we design this environment? If you can tell me or give me a framework like this, that will make it easier…"*

The third participant was also able to explain the assessment framework. He described it to be potentially useful for not just assessments in immersive environments. This participant also recommended having an additional arrow going from "grading criteria" to "learning outcomes". This will in their opinion, signify whether the learning outcomes are met given the performance of students.

In addition to the GBAF 2, the decision-making guidelines for pre-existing applications adoption were provided to the participants as a tool that could be helpful to them in their teaching practices. The participants found it useful and well thought-through. One of the participants added that they would adopt a pre-existing game for assessments even if the log data was not sufficient for measuring all desired ILOs. This participant said they would consider alternative assessment methods such as traditional tests and observation notes, in such cases.

7.4 Applying GBAF 2 to the Design of Assessments

The GBAF 2 is a modified version of the GBAF designed to account for the limitations of the GBAF. The revised framework incorporates all the required components considered necessary for the assessment of learning in immersive environments. To demonstrate its application, the GBAF 2 was utilised in the design of embedded assessments for a VR health and safety (H&S) application and an AR titration application, and the design of external assessments for a DG as reported below.

7.4.1 Case 1: Assessment design for a VR H&S application

The VR LaboSafe game, an H&S simulation game, was used to demonstrate the effectiveness of the GBAF 2 for assessment design. The VR LaboSafe game (https://github.com/PhilippeChan/VRLaboSafeGameDemo) is a training game developed by Chan *et al.* (2021) to train students and professionals on chemical laboratory safety risks. The gameplay has problem-solving characteristics, requiring players to explore a realistic virtual chemical laboratory to find and eliminate safety risks. Unresolved or incorrectly eliminated safety risks could result in accidents that would negatively affect the performance of players. This VR game provides an environment to train students in safety awareness and practices by simulating dangerous scenarios, which cannot be easily replicated in real-life. The VR LaboSafe game was designed following sound instructional design principles with careful considerations for learning information, motivational elements and VR-induced simulator sickness symptoms (Chan *et al.*, 2021).

Embedded and external assessments were designed and integrated into the game to allow for the automatic grading of students. The VR LaboSafe game was used to teach and assess seven chemical engineering students who were in their 2nd year of study at Newcastle University. They were six male and one female student who had a general awareness of laboratory safety rules from previous laboratory sessions but had not been assessed on the subject. For this case study, the students were divided into two groups due to the limited number of head-mounted devices (HMDs) available. Each session lasted 2 hours and the procedure for each was the same: a description of the goal of the session, a declaration of the potential risks associated with the use of VR HMDs and the request for students to sign the consent forms before going ahead with the VR gameplay. The Meta Quest 2 HMDs, also known as Oculus Quest 2, with hand controllers, were used for the study. The Meta Quest 2 HMDs are some of the most affordable yet sophisticated HMDs available in the market. Pre-training activities were integrated into the gameplay to enable students to familiarise themselves with the devices and the tasks in the VR game. The pre-training included demonstrations on how to grip objects, open doors, teleport, and put on personal protective equipment (PPE). Ethical approval for this study was obtained from the ethics committee body at Newcastle University.

As previously mentioned, the assessment design for the VR LaboSafe game happened during the development of the game. The GBAF 2 was used to design the assessment and to allow for automatic scoring and grading of students during gameplay. The ILOs that are outlined subsequently, informed the game design including the tasks and game metrics in the VR LaboSafe environment. Adopting the GBAF 2, the assessment design components of the VR LaboSafe game are shown in figure 7.3.

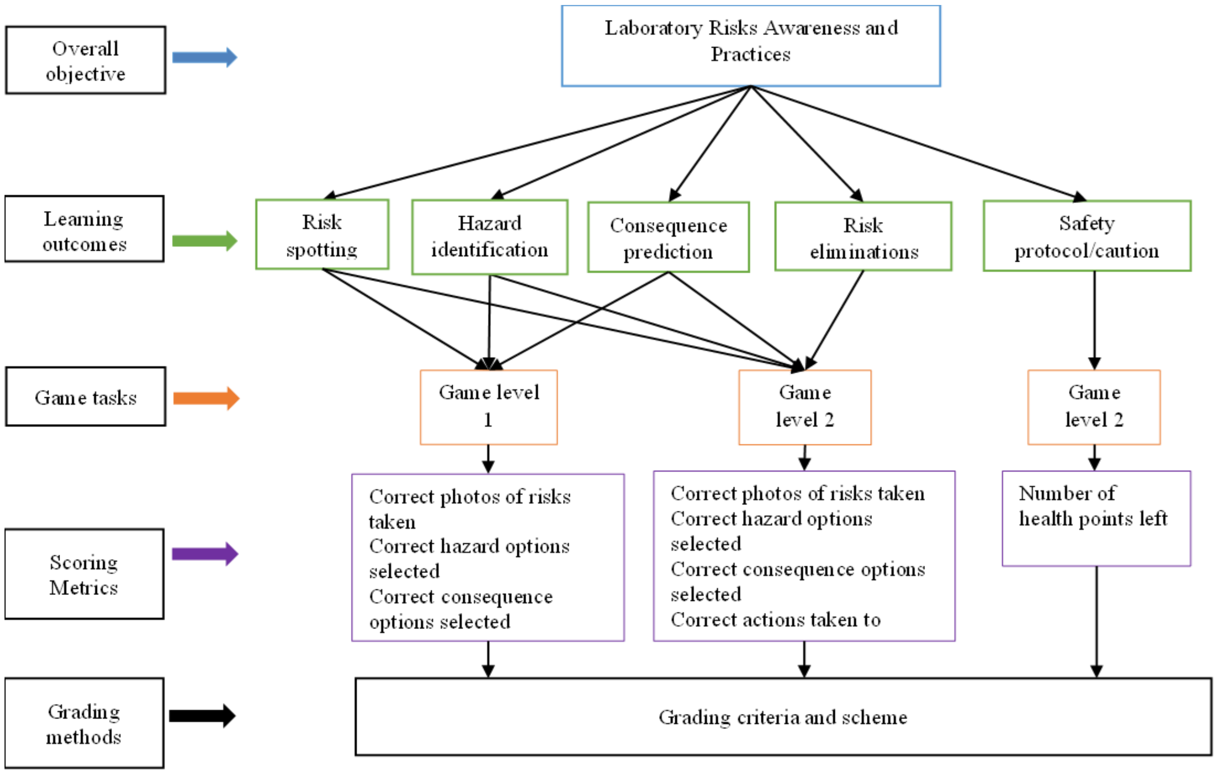


Figure 7.3: Assessment design for the VR LaboSafe game based on the GBAF 2 (Health point is described below).

Overall objective: this VR game aimed to provide an interactive and authentic learning context for students to learn about laboratory risks in order to improve their knowledge and skills in laboratory safety practices.

Learning outcomes: the subject of interest in the game was laboratory risk awareness and practices. To measure the proficiencies of students on these, these competencies were broken down into measurable and achievable ILOs. The ILOs informed the design of appropriate game tasks. The intention was to measure higher-order cognitive process dimensions as outlined below. The cognitive process level and knowledge dimensions following Bloom's taxonomy are also respectively indicated:

1. Students should be able to distinguish hazardous conditions from non-hazardous conditions in the laboratory (Risk spotting) –*Analyse; Factual*
2. Students should be able to evaluate the hazard types associated with each condition identified (hazard identification) –*Evaluate; Conceptual*
3. Students should be able to infer the consequences of laboratory hazards (consequence prediction) –*Understand; Conceptual*
4. Students should be able to apply measures to minimise or eliminate laboratory safety risks (risk elimination) –*Apply; Procedural*

5. Students should be able to safely execute laboratory tasks to minimise risks to themselves (safety protocols) –*Apply; Procedural*

Game tasks: the VR game at the time of this study had two levels that teach and assess laboratory risk awareness and practices. In the first level of the game, students had to search and correctly spot five safety risks in a chemical laboratory. There were always more than five risks in the laboratory and these were randomly presented to students. To complete the task, students were required to take pictures of the identified risky scenario using the virtual tablet provided. Examples of the safety risks presented include a flammable chemical product in a non-explosion-proof fridge, a laboratory technician performing hazardous experiments in a fume hood where the sash was fully open, and a laboratory technician performing an experiment without the appropriate PPE. After spotting the risks, students were presented with multiple choice questions (MCQs) such as, “Which hazard types are present in this safety risk?” and “What are the possible consequences of this risk?” The full list of MCQs is provided in Appendix 7. A. Answering these questions correctly would infer knowledge of hazard types and consequences. In the second level of the game, students were presented with similar tasks as in the first level. Additionally, students were required to eliminate the risks identified. This could mean moving the flammable products to an ignition-free fridge or dressing up a laboratory technician in appropriate PPE. Students could choose to skip this risk-elimination step if they were unable to find a solution, however, this would affect their scores. The game tasks were completed when all five safety risks were spotted in the first level and when all safety risks were (correctly or incorrectly) minimised or skipped in the second level.

Scoring metrics: the scoring metrics selected to assess the ILOs were seamlessly woven into the game tasks in such a way that the gameplay activities of students provided evidence of their competencies. For risk spotting, ILO1, the percentage of correct photos taken within the given time was the scoring metric utilised. For ILO2, hazard identification, students were expected to evaluate the spotted hazard and determine what kind of hazard it posed – chemical, physical, environmental, ergonomic or health hazard. Therefore, to assess this ILO, the percentage of correct hazard options selected from a list of possible options was the scoring metric used. In addition to identifying the hazard types, students were required to infer the potential consequences of such risks. Similar to ILO2, for ILO3, the percentage of correct consequences selected from the list of potential options was used as the scoring metric. Furthermore, to mitigate potential risks in the laboratory, students were expected to make changes to the environment where needed. This involved moving objects, closing fume hoods, or dressing up technicians in the right PPE. For ILO4, students were assessed based on the percentage of correctly mitigated risks. For ILOs 1, 2, 3 and 4, incorrect answers and actions of students were also taken into account as shown in table 7.2. This was done to account for guesswork from students and to evaluate the accuracy of actions.

Lastly, as would happen in a real-world laboratory, interaction with dangerous chemicals could be unsafe without the right PPE. To simulate this effect in the VR game, Health Point (HP) system was incorporated into the second level of the game. The HP of students at the beginning of the game was 5 (100%) but this value was reduced with every inappropriate exposure to dangerous chemicals when appropriate PPE was not worn. Hence, to measure the observance of safety protocols in the laboratory environment, the percentage of HP left after the tasks were determined. The scoring of HP is important as it mimics real-life consequences of poor laboratory practices without causing any harm to students. This ILO was only assessed in the second level of the game as it was previously found that the behaviours of students at the beginning of gameplay are exploratory and prone to mistakes compared to their behaviours subsequently (Udeozor, Russo Abegão, *et al.*, 2021). Assessing the observance of safety protocols with HP in the second level is expected to provide better evidence of the competencies of students on this ILO compared to the first level of the game where students are introduced to the environment and to the subject or topic assessed, for the first time.

Grading methods: with all other components of the GBAF 2 outlined, the grading criteria and scheme were drafted. As this assessment was embedded during the design of the VR game, these formulas for calculating and grading the performance of students were also incorporated. Doing so made it possible to automatically compute and present performance scores to students during gameplay. The scoring methods used to calculate performance on each measured ILO are presented in table 7.2. Given the possibility that students might take random pictures and randomly select options in the MCQs presented in the simulated laboratory environment, the grading methods were designed to penalise wrong actions and answers.

Table 7.2: Grading methods for each assessed ILO in the VR LaboSafe game.

Learning outcomes	**Scoring methods**	**Description**
ILO 1	$R_{spot} = \frac{c}{c+i}$	*c* is the number of correct photos taken and *i* is the number of incorrect photos taken

ILO2	$R_{hazard} = \frac{h}{H+i}$	h is the number of correct hazards types identified, H is the total number of correct hazard options and i is the number of incorrect options selected
ILO3	$R_{conseq} = \frac{k}{K+i}$	k is the number of correct consequences chosen, K is the total number of correct options available and i is the number of incorrect options selected
ILO4	$R_{eliminate} = \frac{l}{L+i}$	l is the number of correctly eliminated risks, L is the total number of risks presented and i is the number of incorrect actions taken
ILO5	$R_{HPoint} = \frac{hp}{5}$	hp is the number of health points remaining after completing the level

In addition to grading each ILO, students were also graded to determine their proficiency levels on each level of the game completed. The following equations were used as scoring methods to calculate the performance of students on each game level:

$$\text{Level 1: } R_{spot} \times 0.4 + R_{\text{hazard}} \times 0.4 + R_{conseq} \times 0.2 \qquad \textit{Equation 7.1}$$

$$\text{Level 2: } R_{spot} \times 0.3 + R_{\text{hazard}} \times 0.3 + R_{conseq} \times 0.2 + R_{eliminate} \times 0.1 + R_{HPoint} \times 0.1 \qquad \textit{Equation 7.2}$$

The coefficients in each scoring method (equations) represent the weight attributed to the assessed outcomes. For Level 1 of the game, weightings of 40%, 40% and 20% were given to risk spotting, hazard identification and consequence prediction, respectively. For Level 2, risk spotting, hazard identification, consequence prediction, risk elimination and HP were weighted 30%, 30%, 20%, 10% and 10%, respectively. These weightings were applied specifically for this demonstration taking into consideration the educational levels of the participants and the level of knowledge expected of them at the time. As the participants of this study were 2nd year students with basic laboratory H&S knowledge and limited experience in using VR for learning, it was only fair that risk spotting and hazard identification carried more weight. These sets of tasks required knowledge of basic risks and hazards in the laboratory and required simple clicks to complete given tasks compared to other tasks. The narrow range chosen for “expert” is indicative of the high level of H&S competency required to be deemed an expert in H&S. The weightings could vary depending on the aim of the game, the knowledge levels of the students, and/or the goal of the assessment.

Finally, the grading scheme was drafted for each ILO and the overall performance on each level of the game. The scoring of performance on the ILOs is considered a formative assessment as it provides students with information about their performance on the set learning outcomes. On the other hand, the grading of each level of the game acts as a summative assessment allocating ratings or scores to students indicating their competency levels. The grading scheme used for the VR LaboSafe game is shown in table 7.3.

Table 7.3: Grading scheme for the VR LaboSafe game.

	Performance Rating		
	Novice	***Competent***	***Expert***
Risk spotting	R_{spot}= 0-40%	R_{spot}= 41-80%	R_{spot}= 81-100%
Hazard identification	R_{hazard}= 0-40%	R_{hazard}= 41-80%	R_{hazard}= 81-100%
Consequence Prediction	R_{conseq}= 0-40%	R_{conseq}= 41-80%	R_{conseq}=81-100%
Risk elimination	$R_{eliminate}$ = 0-40%	$R_{eliminate}$ = 41-80%	$R_{eliminate}$ = 81-100%
Safety protocols/caution	R_{HPoint} = 0-40%	R_{HPoint} = 41-80%	R_{HPoint}= 81-100%
Level 1	0-40%	41-80%	81-100%
Level 2	0-40%	41-80%	81-100%

At the end of each level of gameplay, students were presented with summaries of their performance on scoreboards such as that presented in figure 7.4. These showed their performance and grade levels as either Novice, Competent or Expert on all measured ILO as well as on each level of gameplay. This grading followed the criteria outlined in the grading scheme in table 7.3. Categorising performance levels as either expert, novice or competent/intermediate is well established in the medical field for surgical skills-level assessment (Menekse Dalveren & Cagiltay, 2020; Topalli & Cagiltay, 2019). It is considered useful for assessing the proficiencies of surgeons-in-trainings because it provides an objective method of identifying differences in skills level. For this study, this categorisation scheme is adopted as it is an objective and proven method of identifying competency levels of students on laboratory risk awareness and practices.

Mission 1: Risk Spotting

MISSION COMPLETED ★★★

Time: 08:35 Well done! **HSE Expert**

RISK SPOTTING

Risks spotted:	5/5	★★★
Photos taken:	5	★★★
Hints requested:	0	★★★
Average response time:	0:23	★★★

HAZARD IDENTIFICATION

Correct answers:	8	★★★
Incorrect answers:	1	★★★
Average response time:	0:15	★★★

CONSEQUENCE PREDICTION

Correct answers:	6	★★★
Incorrect answers:	2	★★★
Average response time:	0:18	★★★

TOTAL SCORE: 9540

LEADERBOARD

Your rank

002	KU01420	9540

Top 5

001	KU01069	9837
002	KU01420	9540
003	KU01342	8431
004	KU01994	8129
005	KU01007	7502

Average: 7823

Login: KU01420

Figure 7.4: Sample of the scoreboard presented to students at the end of gameplay.

For a detailed insight into the performance of the students, log files were collected and the retrieved data were analysed. Since this assessment was designed and embedded into the game to measure specific learning outcomes, the data collected and stored were semi-structured and contained only relevant information. Data were collected for each student on each level of the game. To analyse the data, the raw .xml files were converted to readable table format (.csv) using Python 3.7. The outputs were tables containing rows of anonymised identities (IDs) of students and columns of scores on all scoring metrics. In addition to the scoring metrics described earlier, additional data that were considered relevant for research into the learning process of students were collected and these included time on task and hints requested. Although relevant to understanding learning in the game, these were not considered appropriate for assessment and grading at this point. Performance on these metrics can be affected by factors such as familiarity with VR devices, in the case of time on task, and learning style when it comes to hints used. At the time of this study when much information about the VR experience of the students and their learning styles was unknown, incorporating these metrics into the assessment could unfairly affect the performances and grades of students. Using the grading methods in table 7.2, the performance of students on all measured outcomes was computed as shown in table 7.4.

From the results, the overall performance of students was better in level 2 compared to level 1. In level 1, 67% and 23% of the students performed at novice and competent levels respectively, while in level 2, 100% of the students performed at competent levels. This suggests that the VR game was effective for the acquisition of knowledge and skills on risk awareness and practices. This outcome is consistent with others that found DGs, VR and AR effective for improving the performance of engineering students (Bolkas *et al.*, 2022; Criollo-C *et al.*, 2021; Rossado Espinoza *et al.*, 2021; Urbina Coronado *et al.*, 2022; Perini, Oliveira, *et al.*, 2018).

To determine how reliable the grading methods were at establishing these results, sensitivity analyses were performed in Microsoft Excel. The weightings on each assessed competence in levels 1 and 2 were altered to observe how these would impact the performance of students based on the scoring criteria presented in table 7.2. Using equations 7.1 and 7.2, but changing their coefficients to correspond with those in tables 7.5 and 7.6, the outcome shows that the weightings used for the scoring of the performance of students on both levels of gameplay are reliable. In level 1, emphases were placed on measuring how well students mastered risk spotting and hazard identification, while the weightings on these were reduced in level 2 to account for more competencies that are equally important to laboratory risk awareness. Changing the weightings of these competencies will result in changes to the categorisation of performance. This shows that educators can adjust the weights according to the emphases of the training provided.

Table 7.4: Performance of students on measured outcomes and levels of VR LaboSafe gameplay.

Level 1								**Level 2**								
ID	***Time on task (secs)***	***Hints***	***Risk-spot***	***Risk-hazard***	***Risk-consequences***	***Overall score***	***Overall Rating****	***Time on task (sec)***	***Hints***	***Risk-spot***	***Risk-hazard***	***Risk-consequences***	***Risk-eliminate***	***Risk-HPoint***	***Overall score***	***Overall Rating****
CE1	1079	5	0.109	0.273	0.5	0.253	**N**	1147	2	0.556	0.167	0.444	1	0.2	0.481	**C**
CE2	1005	2	0.192	0.444	0.5	0.355	**N**	1109	2	0.625	0.429	0.625	1	0.2	0.624	**C**
CE3	1368	1	0.357	0.545	0.857	0.532	**C**	748	1	1	0	0	0.8	0.2	0.5	**C**
CE4	1549	5	0.063	0.375	0.875	0.350	**N**	830	3	0.217	0.5	0.5	1	0.2	0.457	**C**
CE5	337	1	0.455	0.5	0.5	0.482	**C**	677	1	0.5	0.571	0.75	0.8	0.2	0.621	**C**
CE6	1133	1	0.156	0.4	0.8	0.383	**N**	1211	2	0.278	0.5	0.75	0.8	0.2	0.511	**C**

*Overall ratings were presented in three categories – N for novice, C for competent and E for expert.

Table 7.5: Sensitivity analysis for grading method used for level 1 of the VR LaboSafe game- R_{spot} ×0.4 + R_{hazard} ×0.4 + $R_{consequence}$ ×0.2.

	Original weighting	**Alterations to the weightings**			
ID	**40, 40, 20***	**40, 30, 30**	**60, 30, 10**	**50, 40, 10**	**20, 20, 60**
ChemEng1	0.253	0.275	0.247	0.213	0.376
ChemEng2	0.355	0.360	0.349	0.324	0.427
ChemEng3	0.532	0.564	0.549	0.482	0.695
ChemEng4	0.350	0.400	0.325	0.270	0.613
ChemEng5	0.482	0.482	0.523	0.477	0.491
ChemEng6	0.383	0.423	0.374	0.318	0.591
ChemEng1	N	N	N	N	N
ChemEng2	N	N	N	N	C
ChemEng3	C	C	C	C	C
ChemEng4	N	C	N	N	C
ChemEng5	C	C	C	C	C
ChemEng6	N	C	N	N	C

*Highlighted column shows performance outcomes using the scoring formula for level 1. Other columns show different alterations made to ascertain the sensitivity of the weightings.

Table 7.6: Sensitivity analysis for grading method used for level 2 of the VR LaboSafe game- R_{spot} ×0.3 + R_{hazard} ×0.3 +R_{conseq} ×0.2 + $R_{eliminate}$ ×0.1 + R_{HPoint} ×0.1.

	Original weightings	**Alterations to the weightings**			
ID	**30,30,20,10,10***	**30,30,30,5,5**	**50,20,10,10,10**	**40,40,10,5,5**	**20,20,10,30,20**
ChemEng1	0.481	0.410	0.476	0.393	0.573
ChemEng2	0.624	0.564	0.581	0.544	0.676
ChemEng3	0.500	0.350	0.600	0.450	0.480
ChemEng4	0.457	0.425	0.379	0.397	0.583
ChemEng5	0.621	0.596	0.539	0.554	0.644
ChemEng6	0.511	0.508	0.414	0.436	0.586
ChemEng1	C	C	C	N	C
ChemEng2	C	C	C	C	C
ChemEng3	C	N	C	C	C
ChemEng4	C	C	N	N	C
ChemEng5	C	C	C	C	C
ChemEng6	C	C	C	C	C

*Highlighted column shows performance outcomes using the scoring formula for level 2. Other columns show different alterations made to ascertain the sensitivity of the weightings

To summarise, the overall performance of students on the subject can be said to have generally improved for those students that spent more time in level 1 (which had fewer tasks) compared to those of the other students. These sets of students spent less time completing level 2 of the game and recorded significant improvements in their performance in level 2 compared to level 1 of the game. This may suggest that more extensive engagement with the game potentially led to deeper knowledge and skills development. Similar inferences can be made for students who requested more hints in level 1 of the game. The highest improvement in the performance of students was seen in risk spotting with over 100% improvement in the scores of most of the students. However, for hazard identification and consequence prediction, 50% and 67% of the students performed worse in level 2, respectively. This unexpectedly poor performance could be due to a lack of conceptual understanding of laboratory hazard types and their effects. However, the small sample size of our study limits the conclusions that can be drawn based on these findings. Immersive technologies have been found to have the highest influence on procedural and factual knowledge compared to conceptual knowledge *(*Perini, Luglietti, *et al.*, 2018b; Perini, Oliveira, *et al.*, 2018), which could be the reason for these outcomes. Nonetheless, the overall performance of students on the subject can be said to have generally improved for those students that spent more time in level 1 (which had fewer tasks) compared to the time spent completing tasks in level 2.

7.4.2 Case 2: Assessment design for a mobile AR game

The GBAF 2 was also used to design assessments for the MAR lab application, a mobile markerless AR titration game that was designed to teach titration experimental procedures to students. This application requires students to move, select, grab and combine objects in the augmented space. It replicates real-life laboratory practices by including logbooks and graphs in the laboratory space, in addition to the necessary laboratory equipment. The design of the MAR lab game follows an established AR design framework for education-relevant applications that have been found to be useful for learning titration experiments (Domínguez Alfaro *et al.*, 2022)

Similar steps to those presented in subsection 7.4.1 for the assessment of learning in VR were applied to the design of assessments for the MAR lab game. Assessment considerations were made at the game design phase and so game metrics relevant to the assessment of ILOs were incorporated and grading criteria integrated. Whereas a study was carried out to the test VR LaboSafe game presented in subsection 7.4.1, no study was carried out with the MAR lab game, hence no data collection procedures and results will be presented here. The fully developed MAR lab game was ready to be used towards the end of this research project. This made it impossible to plan and execute a study in time for this thesis. Using the GBAF 2, the assessment components were developed, and alignment between these was established as illustrated in figure 7.5.

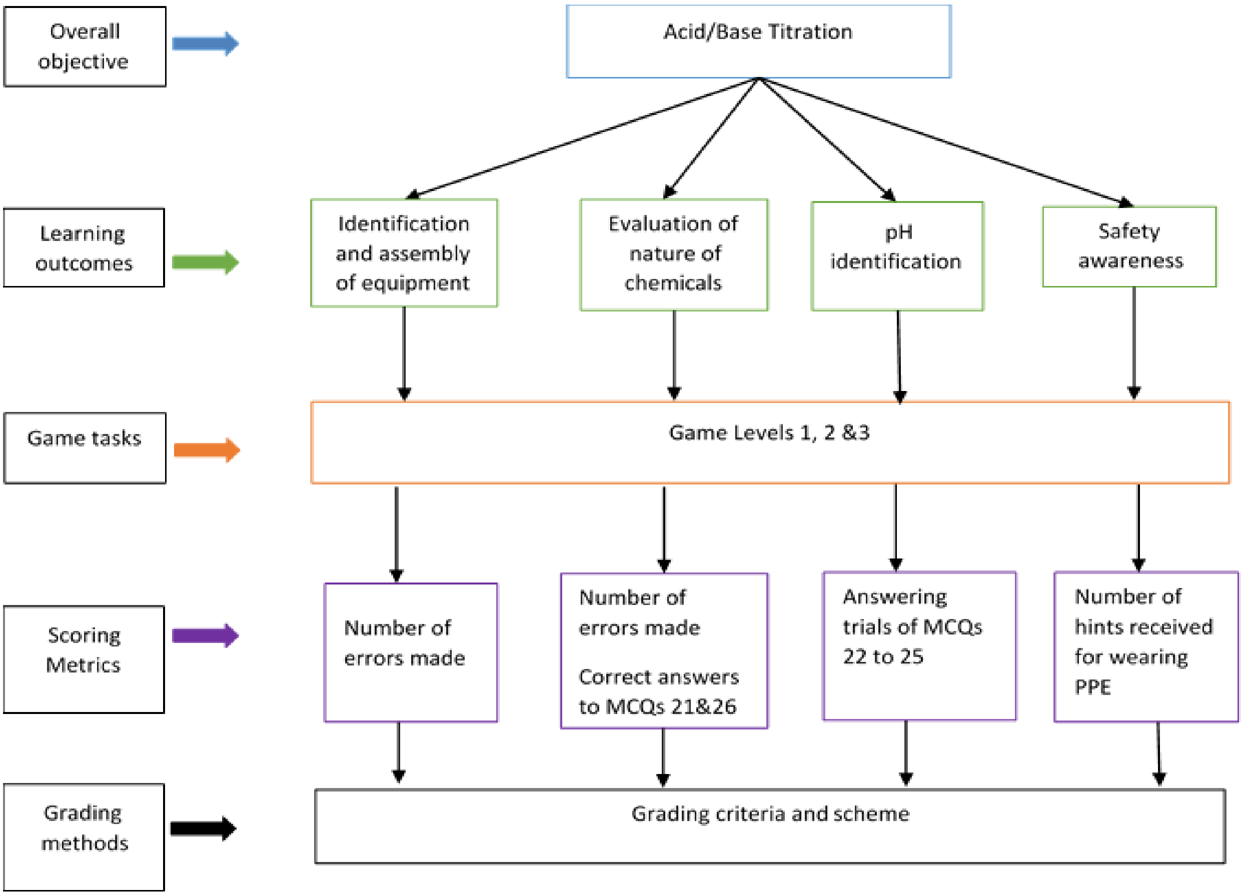


Figure 7.5: Assessment design for the MAR lab game based on the GBAF 2.

Overall objective: the objective was to use the MAR lab game to teach and assess the proficiencies of chemical engineering students on acid-base titration.

Learning outcomes: the assessed ILOs, cognitive process levels and knowledge dimensions according to Bloom's taxonomy (Anderson, 2013) are respectively indicated as follows. During the AR gameplay, students should be able to correctly:

1. Set up the titration equipment – *Apply; Procedural*
2. Assess the nature of the chemicals [acid and base] used for titration - *Evaluate/Understand; Procedural*

3. Calculate the pH of a mixture of acid and base – *Apply; Procedural*
4. Observe safety protocols in the lab – *Apply; Procedural*

***Game tasks*:** the game tasks were divided into levels of gameplay. The tasks for each level of the game are similar but with a slight increase in complexity as players go through the levels. All ILOs are measured in all three levels of the game. The MAR lab game provides a logbook to students at the beginning of gameplay. Students will be required to assemble the titration equipment following the procedure given in the logbook at all levels. This involves correctly assembling the equipment with minimal errors or hints. Safety awareness will also be assessed at all levels as students are expected to put on the appropriate PPE before starting any step in the experiment. In level 1, the students will perform a simple titration (strong acid vs. strong base) and will be introduced to the equipment, the indicator, and the process. In levels 2 and 3, students would need to perform a weak acid vs. strong base titration. At these levels, students will be assessed based on their ability to identify the chemicals correctly. This includes choosing the correct indicator and sorting the waste correctly depending on its nature (acid or base). At all levels, students will be assessed on their ability to identify the pH of the reaction and chemicals used. This would involve selecting the correct answer to the MCQs presented in the game to assess some of the ILOs. It is worth mentioning that the system does not register incorrect answers immediately but allows the students to select another answer until the correct answer is identified; therefore, the answers are registered as trials.

***Scoring metrics*:** the scoring metrics used to assess performance in the MAR game include correct and incorrect actions, errors made, and correct and incorrect answers chosen. To assess ILO1, students are assessed based on the number of errors made while selecting the right equipment and setting up the experiments following the instructions in the logbook. To assess knowledge of the nature of chemical substances, ILO2, the number of errors made when choosing indicators and sorting materials based on their acidic/basic properties, and the number of trials made when choosing correct answers the two MCQs are used as scoring metrics. To assess proficiency in measuring the pH of substances, four MCQs are used (see Appendix 7. B). Lastly, safety protocols are assessed throughout the game using the number of hints students received on the use of appropriate PPE.

Grading methods: the performance of students in each level of gameplay is determined using the formula at each level. For instance, for Level 1 of the game, the goal is to get students to learn how to correctly assemble pieces of equipment for titration experiments, hence the 40%

weighting. In subsequent levels, the emphasis moves to the assessment of more complex tasks such as chemical distinction and pH identification.

- At level 1, the students are just familiarising themselves with the application, the materials, and the procedure. Therefore, “assembly of equipment” has a higher weight.

Level 1 = *Assembly of equipment × 0.4 + Distinction of chemicals × 0.2 + pH identification × 0.1 + safety awareness × 0.3* *Equation 7.3*

- In the second level, students are now expected to be familiar with the application and the experiment. Therefore, the emphasis shifts to the choice of chemicals, and on pH identification (the MCQs in this level measure conceptual understanding).

Level 2 = *Assembly of equipment × 0.2 + Distinction of chemicals × 0.3 + pH identification × 0.4 + safety awareness × 0.1* *Equation 7.4*

- In the last level, equal emphasis is put on the distinction of chemicals, and on pH identification (the MCQs in this level focus on the calculation).

Level 3= *Assembly of equipment × 0.1 + Distinction of chemicals × 0.4 + pH identification × 0.4 + safety awareness × 0.1* *Equation 7.5*

To enable the identification of the proficiency levels of students and rate their performance across the three levels of the game, a grading scheme shown in table 7.7 was developed. Three categories of performers were specified. Performance at a Novice level would lead to no higher than a 40% score while the scores are higher for the Competent and Expert levels. As earlier mentioned, data were not collected for the MAR game, therefore, this section only presents the assessment design considerations made.

Table 7.7: Grading scheme for MAR Lab game.

	Performance rating		
	Novice	**Competent**	**Expert**
Level 1	0-40%	41-80%	81-100%
Level 2	0-40%	41-80%	81-100%
Level 3	0-40%	41-80%	81-100%

7.4.3 Case 3: Assessment design for a DG

Lastly, the GBAF 2 was used to design external assessments for the CosmiClean game. As with the assessment design for the MAR lab game, no data on the assessments of learning in the game is presented here. A study was planned but data collection proved difficult. This was mainly because the study was conducted remotely during the peak of the COVID-19 pandemic and students failed to participate. CosmiClean as was previously described in subsection 4.3.4, is an educational game that teaches the principles of recycling operations to players. In Chapter 6, the application of the GBAF to the design of embedded assessment for CosmiClean was demonstrated. Here, the revised framework is used in retrospect to design MCQs for the game. Although an illustration of the assessment design process for a pre-existing game with no student data is presented in this section, it was important to show how the GBAF can be used in retrospect. For most educators, it is only practical to use pre-existing game applications as developing new bespoke games will be challenging. Therefore, a step-by-step process for designing assessments for pre-existing immersive applications is considered most relevant to educators. Hence, this subsection demonstrates how the GBAF 2 can be applied to the design of traditional assessment tasks external to the game environment. As shown in figure 7.6, the overall objective of the DG is to teach and assess the competencies of students on separation operations by physical properties.

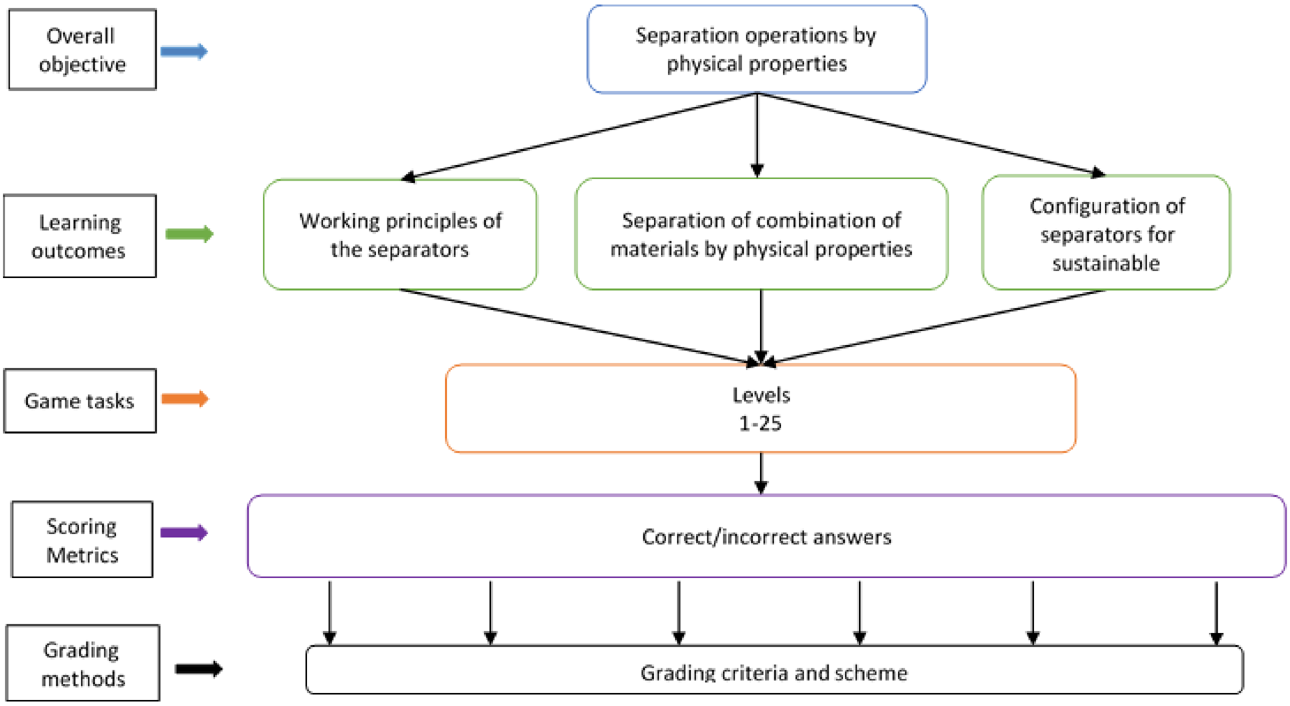


Figure 7.6: Assessment design framework for the CosmiClean based on the GBAF 2.

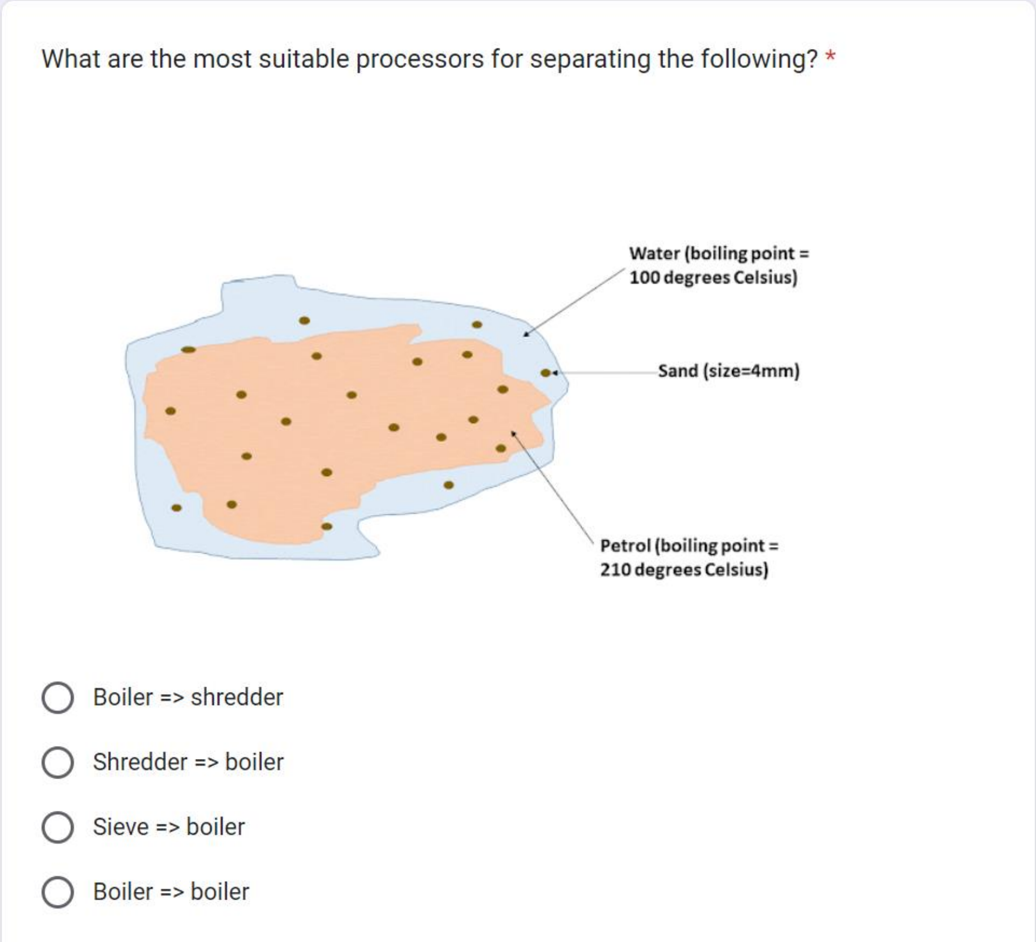
Overall objective: the overall objective of this assessment is to measure the proficiency levels of chemical engineering students in separation operations by physical properties.

Learning outcomes: for this external assessment of learning in CosmiClean, students will be assessed on the three ILOs listed below. The cognitive process levels and knowledge dimensions following Bloom's taxonomy (Anderson, 2013) are also respectively indicated in the ILOs statements. Having completed 25 levels of CosmiClean, students should be able to correctly:

1. Identify what each of the six processors introduced in the game is best used for **-** *Remember; Conceptual*
2. Infer the most suitable processor(s) for separating a combination of particles of different physical properties **-** *Understand; Conceptual*
3. Configure processors to sustainably separate particles of different physical properties **-** *Apply; Procedural*

Game tasks: the tasks for each level of the game are similar but with increasing complexity. Students are required to drag and drop separation equipment, conveyor, and receptors into appropriate positions. The sieve, melter, magnet, shredder, Eddy current separator and stream separators are introduced to the students at different levels of the game in that order. As detailed in section 6.6 of Chapter 6, the processors are introduced to teach and assess the understanding of how the different unit operations work and how to optimise separation by configuring the processors. As these processors are introduced in compounding order such that any introduced processor remains available during gameplay while new ones are introduced, students will also be assessed on their ability to incorporate any of the previously introduced processors to sustainably achieve separation.

Scoring metrics*:* unlike the previous embedded assessments designed using the GBAF, external assessments in the form of MCQs are designed here to assess learning gains in CosmiClean. Educators may on some occasions be unable to carry out embedded assessments or may wish to assess students using familiar methods of assessments. In such cases, pre-game and post-game tests would provide useful information for educators to determine the learning progress made by students. The GBAF 2 can guide the design of other forms of assessments such as MCQs and essay questions. Here, the design of these types of assessments around CosmiClean is presented. Using the GBAF 2, two similar sets of questions that measure the proficiencies of students on the desired ILOs were developed. The goal here is to demonstrate how pre and post-game test items can be developed to assess learning effectiveness when an immersive learning intervention is carried out. The MCQs for this assessment are designed to measure lower and higher-order cognitive processes. MCQs are often limited in their ability to assess higher-order cognitive processes (Villarroel *et al.*, 2019), however, the questions developed to measure the three ILOs outlined are designed to be contextual and performance-based. The questions are designed around the game in such a way that they require students to apply similar knowledge and skills used during the gameplay towards answering the MCQs as illustrated in figure 7.7. The full list of the MCQs can be found in Appendix 7. C. These test items were validated by three academics who are knowledge and pedagogy experts at Newcastle University, UK.



What are the most suitable processors for separating the following? *

○ Boiler => shredder

○ Shredder => boiler

○ Sieve => boiler

○ Boiler => boiler

Figure 7.7: Sample of an MCQ designed for the assessment of learning gains in CosmiClean.

Grading methods: as with traditional MCQs, the scoring metrics to be used are the number of correct and incorrect answers provided by students. The scores awarded to students can be in percentages or categorical scales, as was used for the embedded assessments in the previous sections. Table 7.8 present a grading scheme that can be used for this purpose.

Table 7.8: Grading scheme for scoring the design external assessments for CosmiClean.

	Performance rating		
	Poor	**Good**	**Excellent**
ILO 1	Answered less than 40% of the questions correctly	Answered 41-80% of the questions correctly	Answered 80% or more of the questions correctly
ILO 2	Answered less than 40% of the questions correctly	Answered 41-80% of the questions correctly	Answered 80% or more of the questions correctly
ILO 3	Answered less than 40% of the questions correctly	Answered 41-80% of the questions correctly	Answered 80% or more of the questions correctly

7.5 Assessment Best Practices for Immersive Learning

The GBAF 2 provides a useful guide for the design and implementation of assessments given its successful applications to the design of different assessment types for different environments. While the assessment design considerations proposed by this framework offer important guidance to educators, a description of best practices for the implementation of assessment with immersive learning technologies will provide wholesome guidelines to HE educators. Based on the findings of the preceding chapters, the following are considered best practices when assessing learning gains in immersive learning environments:

1. A chosen immersive learning environment must be engaging. As was found in Chapter 3 of this thesis, students want elements of fun in game-based learning environments. It is not enough that immersive learning environments replicate reality; they should provide a lot more opportunities to experiment and manipulate variables (Dodds, 2021). Tutorials and adaptive interactions in immersive learning environments are recommended as scaffolding measures that reduce the cognitive loads of students and promote engagement (Kalyuga & Plass, 2009; Desurvire *et al.*, 2004). A poorly designed educational game or simulation might not only fail to engage students in the learning activities but could result in invalid outcomes of assessments.
2. Contents of any immersive learning application identified for students must be aligned with curriculum outcomes. Students, particularly adult learners, expect that any innovative learning technology given to them should be relevant to their learning goals. Carrying out assessments in immersive environments that are not well-linked to the curriculum, will result in a waste of time for both students and the educator.

3. Immersive applications should be used for formative rather than summative assessments. These environments provide excellent opportunities to assess and improve the learning performance of students using real-time process data of their interactions in the immersive learning environment. The use of immersive learning applications for summative assessments should only be considered after students have become comfortable with the technology and the application.
4. Allow plenty of gameplay time before assessing learning gains in a new immersive environment. The interactions and behaviours of students in unfamiliar immersive learning environments are erratic or exploratory at the beginning, with more strategic behaviours observed afterwards. Allowing some time before assessment ensures that valid inferences are made about the proficiencies of students. It also minimises the resistance of students toward a new form of assessment as it is often the case that changes to the ways things are normally done are not always welcomed (Åkerlind & Trevitt, 1999). When extended gameplay time poses some challenges to assessment, such that the line between conceptual knowledge and gameplay dexterity is blurred, additional assessment methods such as traditional external assessments should be considered.
5. Gauge the experiences of students with any immersive technology when considering its use for assessment. Based on the findings of this research, the gameplay experiences of students might have some influences on their performance in educational games. Understanding the experiences of students, and accounting for the differences in experience by ensuring that all students reach similar levels of confidence and comfort using the technology before assessments, would enhance the validity, reliability and fairness of the assessment items and outcomes.
6. Immersive technologies, particularly VR, may not be suitable for every learner because of the risks of triggering motion sickness and epilepsy in some learners. Alternative assessment arrangements should be made to accommodate every learner when considering measuring learning gains in immersive environments. This could be the use of traditional assessment methods or through group assessments with immersive technologies, where students can work in groups in different capacities to achieve the set tasks.
7. For a valid assessment design when using immersive learning applications, an established framework or methodology should be used. The GBAF 2 provides a user-

friendly alternative to the ECD framework for educators designing game-based or simulation-based assessments.

7.6 Limitations

One limitation of this chapter is the small sample size of the educators who evaluated the GBAF 2. This chapter could have benefitted from a quantitative evaluation of the framework to demonstrate its usability. Furthermore, the educators who took part in the evaluation had little to no experience with immersive technologies and none had prior experience using these for teaching. The lack of experience with immersive technologies limited the quality and quantity of feedback that these educators could offer. Furthermore, due to time limitations, no validations were carried out on the proposed best practices which would have strengthened the recommendations made. Lastly, a comparative evaluation of the outcomes of the assessments designed using the GBAF with other validated assessment tasks would have further strengthened the argument made, but this is missing in this chapter.

7.7 Conclusions

The increase in the adoption of immersive learning technologies in HE makes the need for an educator-friendly assessment design framework pertinent. The GBAF 2 presented in this chapter is an improvement over the original GBAF introduced in Chapter 6. The uncoupling of step 3 of the GBAF from the other design steps of the original GBAF was necessitated by feedback from educators and other limitations observed during its application to the design of an assessment for an educational DG. The GBAF 2 focuses on highlighting the factors essential for designing assessments that can be used to measure learning in non-traditional complex learning environments. The GBAF 2 is simpler to understand than GBAF and it accounts for all relevant components for the assessment of learning in DGs, VR and AR applications. The GBAF 2 shows the necessary links that need to be made between different essential components for an effective assessment design. The ease of use of the GBAF 2 is demonstrated by its implementation in the design of assessments for VR, AR and DG applications. It also allows for the design of external MCQs that are contextual and performance-based.

The results of the evaluation of the GBAF 2 showed that it is easy to understand and useful for assessment design. The educators who took part in the evaluation noted that frameworks and decision-making guidelines are needed to help educators apply these innovative technologies to classroom activities. Although the qualitative evaluation of the framework showed that it is easy to understand and apply to assessment designs, this finding is limited by the experiences of the evaluators. This is evident in their emphasis on the accessibility and design of curriculum-relevant applications which is logically the first step towards the adoption of immersive technologies for assessment. Nevertheless, this finding highlights the relevance of this project as it is helping educators to think about the assessment implications of these technologies at the very beginning.

7.8 Future Studies

Whereas some constructive feedback was received from the evaluators, more concrete criticism or feedback on the GBAF 2 would have come from educators currently using DGs, VR or AR applications. It would also be useful to have those educators apply the framework to the design of an assessment for an immersive learning environment. Future studies should consider evaluating the GBAF 2 with educators who have some experience with using these technologies for teaching and assessment. Additionally, quantitative data should be collected to provide evidence of the usability of the framework for assessment designs.

In this chapter, a qualitative evaluation of the GBAF 2 was carried out, in addition to its application to all three immersive learning environments of interest. These applications of the GBAF 2 show that it is robust for the design of both embedded and external assessments. While this chapter demonstrates that the GBAF 2 is easy to use and offers a new approach to the design of assessments for immersive learning, a wider application of this framework to assessment designs is still needed. Future studies should consider this structured approach to assessment design when designing assessments to measure the learning effectiveness of immersive technologies.

To show how the outcomes of the assessments of students with tasks designed with the GBAF 2 compare with other validated assessment items, a comparative study should be carried out. This will provide validation for the GBAF 2 and further strengthen the argument for its adoption.

7.1 **Note:** Part of this chapter has been submitted for publication to the **Assessment and Evaluation in Higher Education** journal and to the **Immersive Learning Research Network 2023 conference**

Udeozor, C., Chan, P., Abegão, F. R., & Glassey, J. (2023). Assessment Framework for Virtual Reality, Augmented Reality and Digital Game-Based Learning. *International Journal of Educational Technology in Higher Education*

Udeozor, C., Domínguez Alfaro J.L., & Glassey, J. (2023). Assessment framework for immersive learning: application and evaluation. *9th International Conference of the Immersive Learning Research Network.*

Chapter 8. Conclusions and Future Directions

This chapter summarises the outcomes of the research presented in the previous chapters. It details the results, contributions and conclusions of each chapter, as well as the strength and weaknesses of the research designs used. This chapter is concluded with a discussion of the contribution of the research project to the body of knowledge. It also discusses potential future research directions.

8.1 Conclusions

The goal of this research was to develop an assessment framework aimed at guiding educators through designing and implementing assessments when teaching with immersive technologies. The research was focused on the applications of digital games (DGs), virtual reality (VR) and augmented reality (AR) in higher education (HE), specifically in the engineering education context. The assessment framework developed in this research project was intended to be easy to use and useful for the design of assessments of learning in immersive environments. This overarching research goal resulted in six research studies that were presented in Chapters 2 to 7 of this thesis. In Chapter 2, background of the subject under investigation was provided and a survey of the literature was carried out to understand the state-of-the-art of assessments with immersive learning technologies. Chapter 3 examined the factors that influence the adoption of DGs by engineering students. Chapter 4 was borne out of the need to explore the findings made in Chapter 3, hence the relationship between the experiences of students, their perceptions of learning with DGs and their performance in an educational DG was explored. Chapter 5 examined the gameplay behavioural patterns of students in an educational DG. Lastly, Chapters 6 and 7 presented the development, evaluation and applications of an assessment framework to the design of assessments for a DG, a VR and an AR application. These chapters played different roles toward the goal of this PhD research as detailed below.

8.1.1 Assessments of learning in immersive environments: a systematic review of literature in engineering education

At the start of any primary research, an understanding of current trends in the area of interest is often achieved through the survey of existing literature. Chapter 2 of this thesis presented a systematic and extensive literature review on DGs, VR and AR applications in engineering education published between 2012 and 2022. First, an overview of assessment in

HE was presented. DGs, VR and AR were also introduced, and their potential for teaching and assessments was discussed. The systematic review aimed to identify the assessment methods used in the reported studies in order to determine whether a structured assessment methodology for immersive learning exists. Additionally, this chapter outlined the engineering disciplines where these immersive technologies were used, the types of DGs used, and the AR and VR-enabled devices utilised. The experimental research designs adopted and the outcomes of the effectiveness of these technologies for engineering education were also discussed. A total of 66 papers published in the five databases searched were found to meet the pre-determined inclusion criteria. The findings of the review indicated that external assessment methods in the forms of pre- and/or post-intervention tests, delayed exams, and questionnaires were predominantly used in the reviewed papers. No assessment design framework was identified in the reviewed papers. Papers reporting on the use of immersive technologies for mechanical engineering education were more prevalent than for any of the other engineering disciplines identified in the review. Head-mounted devices and mobile phones were the most frequently reported devices in the VR and AR studies, respectively. Quasi-experimental research designs were predominantly used in the identified papers and this was followed by randomised trials and time series experiments. Generally, the outcomes of the studies indicated that immersive technologies are as effective as, and in some cases better than, conventional teaching tools for engineering education. This is based on the comparative measures reported in the identified studies.

The strength of this chapter lies in the thoroughness of the search of the literature on three innovative teaching tools. Until the publication of this thesis, no such structured and extensive review was published on this topic. Five well-established databases were searched for each immersive technology, each search covering papers published in the past decade. This extensive search resulted in thousands of papers for each technology. Each paper was carefully screened by the researcher and relevant papers were analysed. However, this does not guarantee that every relevant paper was included in the review, a general problem with literature reviews. Some papers that did not use the appropriate keywords in titles, abstracts and keywords sections might have been filtered out. Additionally, papers not indexed in the five searched databases would have been excluded. Nonetheless, this chapter provides a broad overview of the current practices with immersive technologies in the engineering education domain.

8.1.2 Perceptions and factors affecting the adoption of digital games by engineering students

Interests in the use of immersive technologies for education are on the rise with educators exploring them for classroom learning and teaching (L&T). It is established that educators have positive views of the efficacy of these technologies and are keen to use them to improve their teaching and the experiences of students. But as non-traditional learning tools, the views of engineering students towards the use of these technologies for learning and assessments are not well established. Chapter 3 of this thesis presented an explanatory sequential mixed-methods research that aimed to identify factors affecting the adoption of educational DGs by engineering students. Using a questionnaire from an adapted model, the extended Unified Theory of Acceptance and Use of Technology (UTAUT 2), the chapter examined which of the five UTAUT 2 constructs influenced the adoption of DGs by engineering students. The analyses of data from 125 chemical engineering students showed that only hedonic motivation (HM) had a significant influence on the behavioural intentions (BIs) of students to use DGs for learning. A follow-up qualitative study conducted as a focus group discussion with seven chemical engineering students showed that fun and enjoyment in DGs (HM construct) were important factors that would positively influence their use of DGs for learning. Additionally, the relevance of a given DG to the curriculum was also raised by students as an important factor in their adoption of educational DGs. This particular finding from the participants of the focus group discussions can be interpreted as performance expectancy (PE) in the UTAUT 2 model. Notwithstanding the small sample size, the findings highlighted the relevance of mixed-methods research given that this factor did not appear to be influential in the quantitative study. Furthermore, the chapter reported that although students were receptive to the use of DGs for learning, students opposed their use for assessments. One of the main reasons given for this was that digital gaming skills were likely to interfere with their academic performance if assessed with DGs. This concern raised questions regarding the influence of prior gameplay experience on performance in educational DGs.

In summary, Chapter 3 provided an in-depth understanding of the perceptions of engineering students towards the use of DGs for learning and assessments, given the advanced and rigorous mixed-methods research methodology adopted. It allowed for a comprehensive exploration of the findings from the quantitative data through group discussions. One drawback of this chapter, however, is the subjectivity associated with qualitative data analysis. This limitation was minimised through the triangulation of findings from both qualitative and quantitative data, thus enhancing the credibility of the research study. Lastly, the focus group interview consisted of a small sample size of students from a single university. This limits the ability to generalise the findings of the qualitative study across the wider population. Future studies should consider purposefully recruiting more students from diverse institutions for more generalisable outcomes.

8.1.3 Relationship between perceptions, experience and performance of students in an educational game

To address some of the concerns raised by students in the preceding chapter, Chapter 4 explored the relationship between the gameplay experiences of engineering students, their perceptions of educational DGs and their performance in an educational DG. An explanatory correlational research design was used to identify and explain relationships between the three variables in order to determine whether prior gaming experiences influenced performance in DGs. Data collected for this analysis were responses to the UTAUT 2 questionnaire used to measure the perceptions of students in Chapter 3. Additional demographic data were collected on the gameplay experiences of students. Log data from the CosmiClean gameplay of students were used to measure their performance in the game. Using a 2-step clustering technique, the student participants were grouped into two performance groups and the characteristics of their performance were highlighted. The results of Spearman’s correlation analyses of the data of 44 chemical engineering students showed no relationship between the perceptions of students and their performance in the DG. A significant positive relationship was, however, found between the performance of students and their reported enjoyment of gameplay. Within the statistical limitations of the data analysed, it was inferred that technology-savvy students and those who enjoy gameplay would likely outperform their counterparts in educational DGs. This implies that educators would need to account for this effect when considering using educational DGs for assessments. This could mean ensuring that all students are comfortable with a given immersive technology and application, before assessing students with it.

With little to no research on this subject, this chapter highlights the need for further research addressing the influence of prior gameplay experiences on the performance of students in immersive environments. Within the statistical limitations of the data collected, the findings reinforce the concerns raised by students and the need for further investigations.

The drawback of this chapter lies in the small sample size which limits the interpretations of its findings. Although the sample size used in this chapter was sufficient for the analysis conducted, in the future, larger sample sizes should be considered to strengthen the arguments based on the data obtained.

8.1.4 Gameplay behavioural patterns and performance of engineering students in an educational digital game

Chapter 4 highlighted the need to further explore the issue of gameplay experience interferences with performance in educational DGs. In Chapter 5, therefore, the gameplay behaviours of students and how these could affect outcomes of assessments, is investigated. A grounded theory research design was adopted to explore the behavioural patterns of engineering students in an educational DG to provide insights into best practices for assessment designs and implementations. Additionally, the performance of students in the game was assessed by their gameplay log data using a 2-step clustering technique to provide an understanding of their solution patterns. The highlight of this chapter was the identification of two sequential behavioural patterns in the gameplay data of students: exploratory and strategic gameplay behaviours. The gameplay data of 58 chemical engineering students showed multiple unique but inefficient solutions at the start of gameplay, and more strategic solutions were observed subsequently. The cluster analyses were used to identify performance groups of students and the solution characteristics of each group. The analyses of the data of 32 chemical engineering students showed the solution trends across the three identified performance groups.

While the outcomes of this chapter must be cautiously interpreted given the limitation posed by the sample size, the outcomes have considerable implications on assessments in DGs or any other immersive learning environment. Assessing learning or the performance of students after only a short duration of gameplay would likely produce misleading results. As recommended in Chapter 8, the integration of tutorials and adaptivity into immersive learning environments can promote engagement and reduce the cognitive load of students, making it easier for students to familiarise themselves with the learning environment (Desurvire *et al.*, 2004; Kalyuga & Plass, 2009). Choosing a well-designed educational application that incorporates these in the learning environment, might shorten the exploratory phase of the gameplay, allowing for assessments to be carried out sooner. Alternatively, considerable gameplay times should be allowed before assessments are carried out in any immersive learning environments. This is to ensure validity, reliability and fairness of the assessment process. However, in some cases, embedded assessments alone may not be sufficient to measure the in-depth conceptual understanding as the performance of students on the skills dimension may not necessarily mean conceptual understanding (Crawford & Colt, 2004). Where extensive gameplay time is allowed and the measurement of conceptual knowledge necessary, multimodal assessment methods should be considered. Another drawback of this study is the subjectivity associated with grounded theory research design, and the choice of game metrics analysed. Although subjective interpretations of the results of the sequential behavioural pattern analysis were provided in this chapter, quantitative data were analysed to reach logical conclusions. Nonetheless, there is a need for future studies on this to determine whether the patterns identified in this chapter are observed in other educational DGs, especially when other metrics are used for analyses.

8.1.5 A Game-based assessment framework for immersive learning environments

Given the findings of the previous chapters, Chapter 6 proposed an assessment framework that would guide educators to design and implement assessments when using DGs, VR and AR applications in the classroom. A design-based research methodology was used in this chapter. The chapter described the design of the game-based assessment framework (GBAF) building on established assessment methodologies: the Constructive Alignment principles and the Evidence-Centred Design (ECD) framework. The GBAF provided an educator-friendly assessment framework relevant for assessment design and implementation when teaching with immersive learning applications. The GBAF outlined the necessary steps that should be taken by educators to design assessments for the measurement of learning in immersive environments. Additionally, a set of decision-making guidelines was also presented in this chapter to guide educators toward the identification and selection of appropriate pre-existing immersive learning applications to use when it is impossible or impractical to design a new immersive environment. Initial evaluations of the GBAF showed that it can inform assessment designs. Nonetheless, it was found to be complex to understand and missed out on a valuable component for assessments in immersive environments.

The highlight of this thesis is the development of an assessment framework that can guide educators through assessment design and implementation processes. Although this chapter provided a viable assessment framework, the framework had some limitations that necessitated its revision.

8.1.6 The revised game based assessment framework: evaluation and applications

The outcome of the evaluation of the GBAF in Chapter 6 warranted some modifications to the framework. In Chapter 7, the original framework was modified to emphasise the components necessary for assessments of learning in immersive environments, as well as the interrelationships between the outlined components. The revised assessment framework (GBAF 2) highlighted the requirements for the overall objectives and the intended learning outcomes of an instruction to be appropriately aligned with the tasks and game metrics within the immersive learning environments of choice. Additionally, scoring criteria and grading schemes should be formulated to ensure consistency throughout the grading process. This revised framework appeared easier to understand and apply. This view was supported by the outcomes of the evaluation of the framework carried out by three educators. They found the GBAF 2 to be easy to understand and follow. They thought it was a handy guide that can be applied to any assessment design, within or outside immersive learning environments. The application of the GBAF 2 to the design of embedded and external assessments for immersive learning applications reported in Chapter 7, showed that it was robust for different immersive learning environments and assessment types. This chapter also outlined seven best practices to follow when assessing students in immersive learning environments.

The GBAF 2 provides a logical guide to educators wishing to use immersive learning technologies for formal classroom teaching. It provides educators with an easy-to-use framework that could also guide their choice of immersive learning environments for classroom learning and teaching. However, one drawback of this chapter is the small sample size of educators recruited for the evaluation of the framework. Future studies should consider carrying out more extensive evaluations with more educators, and preferably, educators with some experience using immersive technologies for teaching.

8.2 Original Contribution to Knowledge

The author believes that this research offers original contributions to the body of knowledge in the field of education. First, it utilised structured research methodologies to understand how students interact in and view the use of immersive technologies for education. It also provided evidence-based insight into the potential influence of gaming experiences on performance, as well as the emergent sequential behaviours of students in immersive learning environments. The valuable pieces of evidence put forward by this research provide solid foundations for future studies. Secondly, this research proposes a novel conceptual assessment framework that is specifically designed to meet the needs of educators interested in the use of immersive technologies for classroom teaching. Thirdly, this research highlights best practices - important points to be considered when implementing assessments in immersive environments, based on findings from the empirical studies conducted. Fourthly, this research provides decision-making guidelines that could guide educators towards the identification and selection of appropriate pre-existing immersive learning applications when it is impossible to design an entirely new application. The last contribution of this research is the step-by-step description of the applications of the proposed framework to the design of both external and embedded assessments for a DG, VR and AR application.

8.3 Future Research Directions

In the development, evaluation and application of the proposed assessment framework, some areas of improvement and future research were identified. The following future research directions are proposed:

8.3.1 Evaluation of the framework

In this research, the proposed assessment framework was evaluated by educators with limited knowledge and experience in teaching with immersive technologies. As the field develops and more educators gain experience with these technologies, further evaluation of the framework by educators with some experience in using immersive technologies would be needed. The research should consider a usability study to determine whether educators can practically use the framework to design assessments for measuring learning in immersive environments. Surveys and interview discussions with educators could be used to evaluate the usability and usefulness of the framework, and to identify any challenges educators might have with its use.

8.3.2 Embedding assessments into immersive environments

This research covered the development of an assessment framework for educators and the applications of the framework to the design of different assessment types. One area that was not addressed in this research is the technical requirements for embedding assessments into immersive environments. An interesting direction for future research would be to address this by presenting steps that can be taken to embed assessments into DGs, VR and AR games.

The programming considerations to note and possibly, the lines of codes that can be modified for assessment integration would be useful to educators.

8.3.3 Effectiveness of the framework

The proposed framework was successfully applied to the design of assessments, however, claims on its effectiveness cannot be made until a comprehensive comparative study is carried out. Nonetheless, previous studies have found comparative outcomes in the performance of students on multiple-choice tests and game-based assessments (Sanchez *et al.*, 2022) indicating the efficacy of games and simulation-based assessments for the measurement of learning. Further research is needed to determine whether similar outcomes are obtained when students are assessed with tasks designed using the GBAF 2 compared to other traditional validated assessment items.

8.3.4 Feedback for immersive learning

Feedback is integral to learning and crucial for learning in complex immersive environments. This research did not extensively discuss feedback implications for immersive learning, but future research is needed to address this gap. An understanding of effective ways of integrating assessment feedback into immersive learning is needed. The quantity and quality of feedback that would make the most impact on learning would also need to be addressed.

References

Adams, D.M., Pilegard, C. & Mayer, R.E. (2016) ‘Evaluating the Cognitive Consequences of Playing Portal for a Short Duration’, *Journal of Educational Computing Research,* 54(2), pp. 173–195.

Akbulut, A., Catal, C. & Yıldız, B. (2018) ‘On the effectiveness of virtual reality in the education of software engineering’, *Computer Applications in Engineering Education,* 26(4), pp. 918–927.

Åkerlind, G.S. & Trevitt, A.C. (1999) ‘Enhancing Self-Directed Learning through Educational Technology: When Students Resist the Change’, *Innovations in Education & Training International,* 36(2), pp. 96–105.

Alhalabi, W. (2016) ‘Virtual reality systems enhance students’ achievements in engineering education’, *Behaviour & Information Technology,* 35(11), pp. 919–925.

Ali, L. (2018) ‘The Design of Curriculum, Assessment and Evaluation in Higher Education with Constructive Alignment’, *Journal of Education and e-Learning Research,* 5(1), pp. 72–78.

Allen, J.W.P. & Bickhard, M.H. (2022) ‘Emergent Constructivism: Theoretical and Methodological Considerations’, *Human Development,*

Almond, R.G. (2015) ‘Tips and Tricks for Building Bayesian Networks for Scoring Game-Based Assessments’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* [Online]. Springer Verlag. pp. 250–263.

Almond, R.G., Kim, Y.J., Velasquez, G. & Shute, V.J. (2014) ‘How Task Features Impact Evidence From Assessments Embedded in Simulations and Games’, *Measurement: Interdisciplinary Research and Perspectives,* 12(1–2), pp. 1–33.

Almond, R.G., Steinberg, L.S. & Mislevy, R.J. (2003) ‘A four-process architecture for assessment delivery, with connections to assessment design’, *University of Illinois,* 1(June), pp. 147–171.

Alpay, E. (2013) ‘Student attraction to engineering through flexibility and breadth in the curriculum’, *European Journal of Engineering Education*, 38(1), pp. 58–69.

Alrashidi, M., Gardner, M. & Callaghan, V. (2017) ‘Evaluating the Use of Pedagogical Virtual Machine with Augmented Reality to Support Learning Embedded Computing Activity’, in *Proceedings of the 9th International Conference on Computer and Automation Engineering - ICCAE ’17*. [Online]. 2017 New York, New York, USA: ACM Press. pp. 44–50.

Alyami, H., Alawami, M., Lyndon, M., Alyami, M., Coomarasamy, C., Henning, M., Hill, A. & Sundram, F. (2019) ‘Impact of Using a 3D Visual Metaphor Serious Game to Teach History-Taking Content to Medical Students: Longitudinal Mixed Methods Pilot Study’, *JMIR Serious Games*, 7(3), p. e13748.

Amorim, J.A. & Tavares de Azevedo, A. (2021) ‘Engineering Education: Towards the Fourth Industrial Revolution’, in Thomas Heinrich Musiolik & Adrian David Cheok (eds.) *Analyzing Future Applications of AI, Sensors, and Robotics in Society*. [Online]. Hershey: IGI Global. pp. 29–46.

Anderson, L. (2013) *A Taxonomy for Learning, Teaching and Assesing: a Revision of Bloom’s Taxonomy.* New York: Pearson.

Anderson, T. & Shattuck, J. (2012) ‘Design-Based Research: A Decade of Progress in Education Research?’, *Educational Researcher*, 41pp. 16–25.

Andreu-Andrés, M.Á. & García-Casas, M. (2011) ‘Perceptions of gaming as experiential learning by engineering students’, *International Journal of Engineering Education*, 27(4 PART II), pp. 795–804.

Anupam, A., Gupta, R., Naeemi, A. & JafariNaimi, N. (2018) ‘Particle in a Box: An Experiential Environment for Learning Introductory Quantum Mechanics’, *IEEE Transactions on Education*, 61(1), pp. 29–37.

Apperley, T.H. (2006) ‘Genre and game studies: Toward a critical approach to video game genres’, *Simulation & Gaming*, 37(1), pp. 6–23.

Ardiny, H. & Khanmirza, E. (2018) ‘The Role of AR and VR Technologies in Education Developments: Opportunities and Challenges’, in *2018 6th RSI International Conference on Robotics and Mechatronics (IcRoM)*. [Online]. 1 October 2018 IEEE. pp. 482–487.

Arieli-Attali, M., Ward, S., Thomas, J., Deonovic, B. & von Davier, A.A. (2019) ‘The Expanded Evidence-Centered Design (e-ECD) for Learning and Assessment Systems: A Framework for Incorporating Learning Goals and Processes Within Assessment Design’, *Frontiers in Psychology*, 10pp. 1–17.

Ashby, F.G., Isen, A.M. & Turken, U. (1999) ‘A Neuropsychological Theory of Positive Affect and Its Influence on Cognition’. Psychological Review 106 (3).

Ashford-Rowe, K., Herrington, J. & Brown, C. (2014) ‘Establishing the critical elements that determine authentic assessment’, *Assessment & Evaluation in Higher Education*, 39(2), pp. 205–222.

Austin, K., Orcutt, S. & Rosso, J. (2001) ‘How People Learn: introduction to learning theories’. The learning classroom: Theory into practice–a telecourse for teacher education and professional development

Bada, S.O. (2015) ‘Constructivism Learning Theory: A Paradigm for Teaching and Learning’, *IOSR Journal of Research & Method in Education*, 5(46), pp. 66–70.

Bahadoorsingh, S., Dyer, R. & Sharma, C. (2016) ‘Integrating serious games into the engineering curriculum - a game-based learning approach to power systems analysis’, *International Journal of Computational Vision and Robotics*, 6(3), p. 276.

Bahja, M., Kuhail, M.A., Hammad, R., Bahja, M., Kuhail, M.A. & Hammad, R. (2021) ‘Embracing Technological Change in Higher Education’, in L. Waller & S. Waller (eds.) *Higher Education - New Approaches to Accreditation, Digitalization, and Globalization in the Age of Covid*. [Online]. London: IntechOpen.

Bairaktarova, D., Van Den Einde, L. & Bell, J. (2019) ‘Using Digital Sketching and Augmented Reality Mobile Apps to Improve Spatial Visualization in a Freshmen Engineering Course’, in *2019 ASEE Annual Conference & Exposition Proceedings*. [Online]. 2019 ASEE Conferences.

Bakkes, S.C.J., Spronck, P.H.M. & van Lankveld, G. (2012) ‘Player behavioural modelling for video games’, *Entertainment Computing*, 3(3), pp. 71–79.

Bartolo, E. (2017) *Council Recommendation: European Qualification Framework*.

Baruah, B. & Mao, S. (2021) ‘An effective game-based business simulation tool for enhancing entrepreneurial skills among engineering students’, in *2021 19th International Conference on Information Technology Based Higher Education and Training (ITHET)*. [Online]. 4 November 2021 IEEE. pp. 01–06.

Bashabsheh, A.K., Alzoubi, H.H. & Ali, M.Z. (2019) ‘The application of virtual reality technology in architectural pedagogy for building constructions’, *Alexandria Engineering Journal*, 58(2), pp. 713–723.

Bearman, M., Smith, C.D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M. & Neumann, D.L. (2012) ‘Systematic review methodology in higher education’, *Higher Education Research & Development*, 31(5), pp. 625–640.

Beavis, C., Muspratt, S. & Thompson, R. (2015) ‘“Computer games can get your brain working”: student experience and perceptions of digital games in the classroom’, *Learning, Media and Technology*, 40(1), pp. 21–42.

Beavis, C., Rowan, L., Dezuanni, M., McGillivray, C., O’Mara, J., Prestridge, S., Stieler-Hunt, C., Thompson, R. & Zagami, J. (2014) ‘Teachers’ Beliefs about the Possibilities and Limitations of Digital Games in Classrooms’, *E-Learning and Digital Media*, 11(6), pp. 569–581.

Beh, H.J., Rashidi, A., Talei, A. & Lee, Y.S. (2022) ‘Developing engineering students’ capabilities through game-based virtual reality technology for building utility inspection’, *Engineering, Construction and Architectural Management*, 29(7), pp. 2854–

2877.

Behrens, J.T., Mislevy, R.J., Dicerbo, K.E. & Levy, R. (2012) ‘Evidence Centered Design For Learning And Assessment In The Digital World’, in *Technology-Based Assessments for 21st Century Skills*. [Online]. pp. 13–53.

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P. & Berta, R. (2013) ‘Assessment in and of Serious Games: An Overview’, *Advances in Human-Computer Interaction*, 2013pp. 1–11.

Bennett, R.E. (2015) ‘The Changing Nature of Educational Assessment’, *Review of Research in Education*, 39(1), pp. 370–407.

Berelson, B. (1952) *Content analysis in communication research*. Glencoe, IL: Free Press.

Berns, A., Reyes Sánchez, S. & Ruiz Rube, I. (2020) ‘Virtual reality authoring tools for teachers to create novel and immersive learning scenarios’, *ACM International Conference Proceeding Series*, pp. 896–900.

Bhute, V.J., Sengupta, S., Campbell, J., Shah, U. V., Heng, J.Y.Y. & Brechtelsbauer, C. (2022) ‘Effectiveness of a large-scale implementation of hybrid labs for experiential learning at Imperial College London’, *Education for Chemical Engineers*, 39pp. 58–66.

Biggs, J. (2003) ‘Aligning Teaching for Constructing Learning’, *The Higher Education Academy*, 1(4), .

Biggs, J. (1999) *Teaching for Quality Learning at University: What the Student Does*. Buckingham: OpenUniversity Press.

Biggs, J. & Collis, K.F. (1982) *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Biggs, J. & Tang, C. (2010) ‘Applying constructive alignment to outcomes-based teaching and learning’, in *Training material for “quality teaching for learning in higher education” workshop for master trainers, Ministry of Higher Education, Kuala Lumpur*.

[Online]. 2010 pp. 23–25.

Bina, S., Mullins, J.K. & Petter, S. (2021) ‘Examining game-based approaches in human resources recruitment and selection: A literature review and research agenda’, in *Proceedings of the Annual Hawaii International Conference on System Sciences*. [Online]. 2021 pp. 1325–1334.

Bodnar, C., Liauw, M. & Sainio, T. (2021) ‘Digitalisation in chemical engineering education and training’, *Education for Chemical Engineers*, 36pp. 202–203.

Bodnar, C.A., Anastasio, D., Enszer, J.A. & Burkey, D.D. (2016) ‘Engineers at Play: Games as Teaching Tools for Undergraduate Engineering Students’, *Journal of Engineering Education*, 105(1), pp. 147–200.

Bolkas, D., Chiampi, J.D., Fioti, J. & Gaffney, D. (2022) ‘First Assessment Results of Surveying Engineering Labs in Immersive and Interactive Virtual Reality’, *Journal of Surveying Engineering*, 148(1), p. 04021028.

Bolliger, D.U., Mills, D., White, J. & Kohyama, M. (2015) ‘Japanese Students’ Perceptions of Digital Game Use for English-Language Learning in Higher Education’, *Journal of Educational Computing Research*, 53(3), pp. 384–408.

Bourgonjon, J., Valcke, M., Soetaert, R. & Schellens, T. (2010) ‘Students’ perceptions about the use of video games in the classroom’, *Computers and Education*, 54(4), pp. 1145–1156.

Boyle, E., Connolly, T.M. & Hainey, T. (2011) ‘The role of psychology in understanding the impact of computer games’, *Entertainment Computing*, 2(2), pp. 69–74.

Boyle, E.A., Hainey, T., Connolly, T.M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C. & Pereira, J. (2016) ‘An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games’, *Computers and Education*, 94pp. 178–192.

Buckley, J., Colosimo, L., Rebecca, K., McCall, M. & Snow, E. (2021) ‘Game-based assessment for education’, in Stephan Vincent-Lancrin (ed.) *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*. [Online]. Paris: OECD Publishing.

Bullen, M. (1998) ‘Participation and Critical Thinking in Online University Distance Education’, *Journal of Distance Education*, 13(2), pp. 1–32.

Cardin, M.-A., Yixin, J., Yue, H.K.-H. & Haidong, F. (2015) ‘Training Design and Management of Flexible Engineering Systems: An Empirical Study Using Simulation Games’, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(9), pp. 1268–1280.

Carless, D. (2015) *Excellence in university assessment: learning from award-winning practice*. London: Routledge.

Carter, N., Bryant-Lukosius, D., Dicenso, A., Blythe, J. & Neville, A.J. (2014) ‘The use of triangulation in qualitative research’, *Oncology Nursing Forum*, 41(5), pp. 545–547.

Chan, P., Van Gerven, T., Dubois, J.-L. & Bernaerts, K. (2021) ‘Design and Development of a VR Serious Game for Chemical Laboratory Safety’, in F. de Rosa, I. Marfisi Schottman, J. Baalsrud Hauge, F. Bellotti, P. Dondio, & M. Romero (eds.) *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [Online]. Cham: Springer. pp. 23–33.

Chang, C.C., Liang, C., Chou, P.N. & Lin, G.Y. (2017) ‘Is game-based learning better in flow experience and various types of cognitive load than non-game-based learning? Perspective from multimedia and media richness’, *Computers in Human Behavior*, 71pp. 218–227.

Chang, Y., Aziz, E.-S., Esche, S.K. & Chassapis, C. (2012) ‘A GAME-BASED LABORATORY FOR GEAR DESIGN’, *Computer in Education Journal*, 22(1), pp. 21–31.

Chang, Y., Aziz, E.-S., Zhang, Z., Zhang, M. & Esche, S.K. (2016) ‘Evaluation of a video game adaptation for mechanical engineering educational laboratories’, in *2016 IEEE*

Frontiers in Education Conference (FIE). [Online]. 28 October 2016 IEEE. pp. 1–6.

Chaturvedi, S., Yoon, J., McKenzie, R., Katsioloudis, P., Garcia, H. & Ren, S. (2012) 'Implementation and Assessment of a Virtual Reality Experiment in the Undergraduate Themo-fluids Laboratory', in *2012 ASEE Annual Conference & Exposition Proceedings*. [Online]. 2012 ASEE Conferences.

Chaudy, Y. & Connolly, T. (2019) 'Specification and evaluation of an assessment engine for educational games: Integrating learning analytics and providing an assessment authoring tool', *Entertainment Computing,* 30p. 100294.

Chaves, I.D.G., López, J.R.G., Zapata, J.L.G., Robayo, A.L. & Niño, G.R. (2016) 'Process Optimization in Chemical Engineering', in *Process Analysis and Simulation in Chemical Engineering*. 1st edition [Online]. Cham: Springer International Publishing. pp. 343–369.

Checa, D. & Bustillo, A. (2020) 'Advantages and limits of virtual reality in learning processes: Briviesca in the fifteenth century', *Virtual Reality,* 24(1), pp. 151–161.

Chen, L.W., Tsai, J.P., Kao, Y.C. & Wu, Y.X. (2019) 'Investigating the learning performances between sequence- and context-based teaching designs for virtual reality (VR)-based machine tool operation training', *Computer Applications in Engineering Education,* 27(5), pp. 1043–1063.

Cheng, K.-H. & Tsai, C.-C. (2013) 'Affordances of Augmented Reality in Science Learning: Suggestions for Future Research', *Journal of Science Education and Technology*, 22(4), pp. 449–462.

Chittaro, L. & Buttussi, F. (2015) 'Assessing Knowledge Retention of an Immersive Serious Game vs. a Traditional Education Method in Aviation Safety', *IEEE Transactions on Visualization and Computer Graphics*, 21(4), pp. 529–538.

Chomeya, R. (2010) 'Quality of Psychology Test Between Likert Scale 5 and 6 Points', *Journal of Social Sciences*, 6(3), pp. 399–403.

Chon, S.-H., Timmermann, F., Dratsch, T., Schuelper, N., Plum, P., Berlth, F., Datta, R.R., Schramm, C., Haneder, S., Späth, M.R., Dübbers, M., Kleinert, J., Raupach, T., Bruns, C. & Kleinert, R. (2019) ‘Serious Games in Surgical Medical Education: A Virtual Emergency Department as a Tool for Teaching Clinical Reasoning to Medical Students’, *c*, 7(1), .

Cohen, J. (1992) ‘A Power Primer’, *Psychological Bulletin*, pp. 155–159.

Cohen, L., Manion, L. & Morrison, K. (2017) *Research methods in education*. Routledge.

Cohen, L., Manion, L. & Morrison, K. (2007) *Research Methods in Education*. 6th edition. Oxfordshire: Routledge.

Coiffet, P. & Burdea, G. (2003) *Virtual reality technology*. 2nd ed.. Hoboken, N.J.: J. Wiley-Interscience.

Coller, B.D. & Scott, M.J. (2009) ‘Effectiveness of using a video game to teach a course in mechanical engineering’, *Computers and Education*, 53(3), pp. 900–912.

Connolly, T., Stansfield, M. & Hainey, T. (2009) ‘Towards the Development of a Games-Based Learning Evaluation Framework’, in *Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces*. [Online]. IGI Global. pp. 251–273.

Connolly, T.M., Boyle, E.A., MacArthur, E., Hainey, T. & Boyle, J.M. (2012) ‘A systematic literature review of empirical evidence on computer games and serious games’, *Computers & Education*, 59(2), pp. 661–686.

Contero, M., Gomis, J.M., Naya, F., Albert, F. & Martin-Gutierrez, J. (2012) ‘Development of an augmented reality based remedial course to improve the spatial ability of engineering students’, in *2012 Frontiers in Education Conference Proceedings*. [Online]. October 2012 IEEE. pp. 1–5.

Cook-Chennault, K. & Villanueva, I. (2020) ‘Exploring perspectives and experiences of diverse learners’ acceptance of online educational engineering games as learning tools in the classroom’, in *2020 IEEE Frontiers in Education Conference (FIE)*. [Online]. 21 October 2020 IEEE. pp. 1–9.

Cook, D.A. & Beckman, T.J. (2006) ‘Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application’, *The American Journal of Medicine,* 119(2), .

Cooper, B., Thorpe, R. & Alpay, E. (2020) ‘Exploring the benefits of project-based pilot plant experience for chemical engineering undergraduates’, in K. Gravett, N. Yakovchuk, & I. Kinchin (eds.) *Enhancing Student-Centred Teaching in Higher Education*. [Online]. Cham: Palgrave Macmillan. pp. 185–199.

Cooper, G., Park, H., Nasr, Z., Thong, L.P. & Johnson, R. (2019) ‘Using virtual reality in the classroom: preservice teachers’ perceptions of its use as a teaching and learning tool’, *Educational Media International,* 56(1), pp. 1–13.

Crawford, S.W. & Colt, H.G. (2004) ‘Virtual Reality and Written Assessments Are of Potential Value to Determine Knowledge and Skill in Flexible Bronchoscopy’, *Respiration,* 71(3), pp. 269–275.

Creswell, J.W. (2011) *Educational Research*. 4th edition. Boston, MA: Pearson Education.

Criollo-C, S., Abad-Vásquez, D., Martic-Nieto, M., Velásquez-G, F.A., Pérez-Medina, J.-L. & Luján-Mora, S. (2021) ‘Towards a New Learning Experience through a Mobile Application with Augmented Reality in Engineering Education’, *Applied Sciences,* 11(11), p. 4921.

Cronin, P. (1997) *Report on the Application of Virtual Reality Technology to Education*.

Crookall, D. (2010) ‘Serious Games, Debriefing, and Simulation/Gaming as a Discipline’, *Simulation & Gaming,* 41(6), pp. 898–920.

Csikszentmihalyi, M. (1990) *Flow: The Psychology of Optimal Experience*. New York: Harper& Row.

Dalmaijer, E.S., Nord, C.L. & Astle, D.E. (2022) ‘Statistical power for cluster analysis’, *BMC Bioinformatics,* 23(1), pp. 1–28.

Dawson, P. (2017) ‘Assessment rubrics: towards clearer and more replicable design, research and practice’, *Assessment & Evaluation in Higher Education*, 42(3), pp. 347–360.

Deshpande, A.A. & Huang, S.H. (2011) ‘Simulation games in engineering education: A state-of-the-art review’, *Computer Applications in Engineering Education*, 19(3), pp. 399–410.

Desurvire, H., Caplan, M. & Toth, J.A. (2004) ‘Using heuristics to evaluate the playability of games’, in *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI ’04*. [Online]. 2004 New York, New York, USA: ACM Press. p. 1509.

Dib, H. & Adamo-Villani, N. (2014) ‘Serious Sustainability Challenge Game to Promote Teaching and Learning of Building Sustainability’, *Journal of Computing in Civil Engineering*, 28(5), .

DiCerbo, K. (2017) ‘Building the Evidentiary Argument in Game-Based Assessment.’, *Journal of Applied Testing Technology*, 18pp. 7–18.

Dirin, A., Alamäki, A. & Suomala, J. (2019) ‘Gender differences in perceptions of conventional video, virtual reality and augmented reality’, *International Journal of Interactive Mobile Technologies*, 13(6), pp. 93–103.

Dixson, D.D. & Worrell, F.C. (2016) ‘Formative and Summative Assessment in the Classroom’, *Theory into Practice*, 55(2), pp. 153–159.

Dodds, H.E. (2021) ‘Immersive Learning Environments : Designing XR into Higher Education’, in J. E. Stefaniak, S. Conklin, B. Oyarzun, & R. M. Reese (eds.) *A Practitioner’s Guide to Instructional Design in Higher Education*. [Online]. London: EdTech Books.

Dolnicar, S., Grün, B. & Leisch, F. (2016) ‘Increasing sample size compensates for data problems in segmentation studies’, *Journal of Business Research*, 69(2), pp. 992–999.

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), pp. 531–537.

Driessen, E., van der Vleuten, C., Schuwirth, L., van Tartwijk, J. & Vermunt, J. (2005) 'The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study', *Medical Education*, 39(2), pp. 214–220.

Duarte, A.J., Malheiro, B., Silva, M.F., Ferreira, P.D. & Guedes, P.B. (2022) 'Active Learning Strategies for Sustainable Engineering: The Case of the European Project Semester', in Benedita Malheiro & Pedro Fuentes-Durá (eds.) *Handbook of Research on Improving Engineering Education With the European Project Semester*. [Online]. Hershey: IGI Global. pp. 146–164.

Duffull, S.B. & Peterson, A.K. (2020) 'Students' perceptions of playing a serious game intended to enhance therapeutic decision-making in a pharmacy curriculum', *Currents in Pharmacy Teaching and Learning*, 12(11), pp. 1348–1353.

Duin, H., Pourabdollahian, B., Thoben, K.-D. & Taisch, M. (2013) 'On the effectiveness of teaching sustainable global manufacturing with serious gaming', in *2013 International Conference on Engineering, Technology and Innovation (ICE) & IEEE International Technology Management Conference*. [Online]. June 2013 IEEE. pp. 1–8.

Erol, S., Jäger, A., Hold, P., Ott, K. & Sihn, W. (2016) 'Tangible Industry 4.0: A Scenario-Based Approach to Learning for the Future of Production', in *Procedia CIRP*. [Online]. 2016 pp. 13–18.

Estriegana, R., Medina-Merodio, J.-A. & Barchino, R. (2019) 'Student acceptance of virtual laboratory and practical work: An extension of the technology acceptance model', *Computers & Education*, 135pp. 1–14.

Eszergár-Kiss, D. & Caesar, B. (2017) 'Definition of user groups applying Ward's method', *Transportation Research Procedia*, 22(2016), pp. 25–34.

Evangelou, S.M., Stamoulakatou, G. & Xenos, M. (2021) 'A Serious Game for Mobile Phones used in a Software Engineering Course: Usability Evaluation and Educational Effectiveness', in *IEEE Global Engineering Education Conference (EDUCON)*.

[Online]. 21 April 2021 IEEE. pp. 219–225.

Fagan, M., Kilmon, C. & Pandey, V. (2012) ‘Exploring the adoption of a virtual reality simulation: The role of perceived ease of use, perceived usefulness and personal innovativeness’, *Campus-Wide Information Systems*, 29(2), pp. 117–127.

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*, 127pp. 252–266.

Flores, N., Paiva, A.C.R. & Cruz, N. (2020) ‘Teaching Software Engineering Topics Through Pedagogical Game Design Patterns: An Empirical Study’, *Information*, 11(3), p. 153.

Fonseca, D., Martí, N., Redondo, E., Navarro, I. & Sánchez, A. (2014) ‘Relationship between student profile, tool use, participation, and academic performance with the use of Augmented Reality technology for visualized architecture models’, *Computers in Human Behavior*, 31(1), pp. 434–445.

Forina, M., Armanino, C. & Raggio, V. (2002) ‘Clustering with dendrograms on interpretation variables’, *Analytica Chimica Acta*, 454(1), pp. 13–19.

Fornós, S., Udeozor, C., Glassey, J. & Cermak-Sassenrath, D. (2022) ‘The CHEM Jam - how to integrate a game creation event in curriculum-based engineering education’, *Education for Chemical Engineers*, 40pp. 8–16.

Franco-Mariscal, A.J., Oliva-Martínez, J.M. & Almoraima Gil, M.L. (2015) ‘Students’ Perceptions about the Use of Educational Games as a Tool for Teaching the Periodic Table of Elements at the High School Level’, *Journal of Chemical Education*, 92(2), pp. 278–285.

Frank, J.A. & Kapila, V. (2017) ‘Mixed-reality learning environments: Integrating mobile interfaces with laboratory test-beds’, *Computers & Education*, 110pp. 88–104.

de Freitas, S. (2006) ‘Learning in Immersive worlds: A review of game-based learning’, *JISC e-Learning Innov*, 3p. 73.

Fry, H., Ketteridge, S. & Marshall, S. (2003) *A Handbook for Teaching and Learning in Higher Education : Enhancing Academic Practice*. 2nd edition. London: Routledge.

Games, P.A., Keselman, H.J. & Rogan, J.C. (1981) ‘Simultaneous pairwise multiple comparison procedures for means when sample sizes are unequal.’, *Psychological Bulletin,* 90(3), pp. 594–598.

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,* 36pp. 12–23.

Garris, R., Ahlers, R. & Driskell, J.E. (2002) ‘Games, motivation, and learning: A research and practice model’, *Simulation and Gaming,* 33(4), pp. 441–467.

Garzón, J., Pavón, J. & Baldiris, S. (2019) ‘Systematic review and meta-analysis of augmented reality in educational settings’, *Virtual Reality,* 23(4), pp. 447–459.

Gee, J.P. (2003) ‘What video games have to teach us about learning and literacy’, *Computers in Entertainment,* 1(1), .

Glassey, J. & Haile, S. (2012) ‘Sustainability in chemical engineering curriculum’, *International Journal of Sustainability in Higher Education,* 13(4), pp. 354–364.

Glassey, J. & Magalhães, F.D. (2020) ‘Virtual labs – love them or hate them, they are likely to be used more in the future’, *Education for Chemical Engineers,* 33pp. 76–77.

Gogtay, N.J. & Thatte, U.M. (2017) ‘Principles of correlation analysis’, *Journal of Association of Physicians of India,* 65pp. 78–81.

Gold, A.H., Malhotra, A. & Segars, A.H. (2001) ‘Knowledge Management: An Organizational Capabilities Perspective’, *Journal of Management Information Systems,* 18(1), pp. 185–214.

Gómez-Tone, H.C., Martin-Gutierrez, J., Valencia Anci, L. & Mora Luis, C.E. (2020)

‘International Comparative Pilot Study of Spatial Skill Development in Engineering Students through Autonomous Augmented Reality-Based Training’, *Symmetry*, 12(9), p. 1401.

Gopalan, M., Rosinger, K. & Ahn, J. Bin (2020) ‘Use of Quasi-Experimental Research Designs in Education Research: Growth, Promise, and Challenges’, *Review of Research in Education*, 44(1), pp. 218–243.

Gordillo, A., Lopez-Fernandez, D. & Tovar, E. (2022) ‘Comparing the Effectiveness of Video-Based Learning and Game-Based Learning Using Teacher-Authored Video Games for Online Software Engineering Education’, *IEEE Transactions on Education*, pp. 1–9.

Goswami, A. & Dutta, S. (2016) ‘Gender Differences in Technology Usage—A Literature Review’, *Open Journal of Business and Management*, 04(01), pp. 51–59.

Gough, D. (2007) ‘Weight of Evidence: a framework for the appraisal of the quality and relevance of evidence’, *Research Papers in Education*, 22(2), pp. 213–228.

Green, T. (2016) ‘A methodological review of structural equation modelling in higher education research’, *Studies in Higher Education*, 41(12), pp. 2125–2155.

Groff, J.S. (2018) ‘The potentials of game-based environments for integrated, immersive learning data’, *European Journal of Education*, 53(2), pp. 188–201.

Guo, W. (2018) ‘Improving Engineering Education Using Augmented Reality Environment’, in P. Zaphiris & A. Ioannou (eds.) *Learning and Collaboration Technologies. Design, Development and Technological Innovation. LCT 2018. Lecture Notes in Computer Science*. [Online]. Cham: Springer. pp. 233–242.

Guo, W. & Kim, J.H. (2020) ‘How Augmented Reality Influences Student Workload in Engineering Education’, in C. Stephanidis, D. Harris, Wen-Chin Li, D. D. Schmorrow, C. M. Fidopiastis, P. Zaphiris, A. Ioannou, X. Fang, R. Sottilare, & J. Schwarz (eds.) *HCI International 2020 – Late Breaking Papers: Cognition, Learning and Games. HCII 2020. Lecture Notes in Computer Science*. [Online]. Cham: Springer.

Gutiérrez, J.M. & Fernández, M.D.M. (2014) ‘Applying augmented reality in engineering education to improve academic performance & student motivation’, *International Journal of Engineering Education,* 30(3), pp. 1–11.

Ha, I., Yoon, Y. & Choi, M. (2007) ‘Determinants of adoption of mobile games under mobile broadband wireless access environment’, *Information & Management,* 44(3), pp. 276–286.

Hainey, T., Connolly, T.M., Chaudy, Y., Boyle, E., Beeby, R. & Soflano, M. (2015) ‘Assessment Integration in Serious Games’, in *Gamification*. [Online]. IGI Global. pp. 515–540.

Hair, J.F., Hult, G.T.M., Ringle, C. & Sarstedt, M. (2017) *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. 2nd editio. SAGE Publications Inc.

Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P. & Ray, S. (2021) *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook*. Cham: Springer.

Halabi, O. (2020) ‘Immersive virtual reality to enforce teaching in engineering education’, *Multimedia Tools and Applications,* 79(3–4), pp. 2987–3004.

Havig, P., Mcintire, J. & Geiselman, E. (2011) ‘Virtual reality in a cave: Limitations and the need for HMDs?’, in *SPIE 8041, Head- and Helmet-Mounted Displays XVI: Design and Applications, 804107*. [Online]. 2011

Herde, C.N., Lievens, F., Solberg, E.G., Harbaugh, J.L., Strong, M.H. & J. Burkholder, G. (2019) ‘Situational judgment tests as measures of 21st century skills: Evidence across Europe and Latin America’, *Revista de Psicología del Trabajo y de las Organizaciones,* 35(2), pp. 65–74.

Herzog, M. & Katzlinger, E. (2011) ‘Influence of learning styles on the acceptance of game based learning in higher education: Experiences with a role playing simulation game’, *Proceedings of the 5th European Conference on Games-based Learning,* (October), pp. 241–250.

Hettinger, L.J. & Riccio, G.E. (1992) ‘Visually Induced Motion Sickness in Virtual Environments’, *Presence: Teleoperators and Virtual Environments,* 1(3), pp. 306–310.

Horvat, N., Martinec, T., Lukačević, F., Perišić, M.M. & Škec, S. (2022) ‘The potential of immersive virtual reality for representations in design education’, *Virtual Reality,* 1pp. 1–18.

Hou, H.-T. (2012) ‘Exploring the behavioral patterns of learners in an educational massively multiple online role-playing game (MMORPG)’, *Computers & Education,* 58(4), pp. 1225–1233.

Hou, H.-T. (2015) ‘Integrating cluster and sequential analysis to explore learners’ flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration’, *Computers in Human Behavior,* 48pp. 424–435.

Hsu, C.-L. & Lu, H.-P. (2004) ‘Why do people play on-line games? An extended TAM with social influences and flow experience’, *Information & Management,* 41(7), pp. 853–868.

Ibáñez, M.-B. & Delgado-Kloos, C. (2018) ‘Augmented reality for STEM learning: A systematic review’, *Computers & Education,* 123pp. 109–123.

Ifenthaler, D., Eseryel, D. & Ge, X. (2012) ‘Assessment for Game-Based Learning’, in Dirk Ifenthaler, Deniz Eseryel, & Xun Ge (eds.) *Assessment in Game-Based Learning: Foundations, Innovations, and Perspectives.* [Online]. Cham: Springer.

Im, T., An, D., Kwon, O.-Y. & Kim, S.-Y. (2017) ‘A Virtual Reality based Engine Training System - A Prototype Development & Evaluation’, in *Proceedings of the 9th International Conference on Computer Supported Education*. [Online]. 2017 SCITEPRESS - Science and Technology Publications. pp. 262–267.

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

Jaffal, Y. & Wloka, D. (2015) ‘Employing game analytics techniques in the psychometric measurement of game-based assessments with dynamic content’. Journal of E-Learning and Knowledge Society 11 (3) p.pp. 101–115.

Jaksic, N. (2018) ‘A Virtual Reality Course Using EON Reality: Students’ Experiences’, in *2018 ASEE Annual Conference & Exposition Proceedings*. [Online]. 2018 ASEE Conferences.

Jones, R.J. (2017) *A brief history of computer games in the classroom*. [Online] [online]. Available from: https://phys.org/news/2017-06-history-games-classroom.html (Accessed 18 October 2022).

Jonsson, A. & Svingby, G. (2007) ‘The use of scoring rubrics: Reliability, validity and educational consequences’, *Educational Research Review*, 2(2), pp. 130–144.

Kalyuga, S. & Plass, J.L. (2009) ‘Evaluating and Managing Cognitive Load in Games’, in R Ferdig (ed.) *Handbook of Research on Effective Electronic Gaming in Education*. [Online]. London: IGI Global.

Kamińska, D., Zwoliński, G., Wiak, S., Petkovska, L., Cvetkovski, G., Barba, P. Di, Mognaschi, M.E., Haamer, R.E. & Anbarjafari, G. (2021) ‘Virtual Reality-Based Training: Case Study in Mechatronics’, *Technology, Knowledge and Learning*, 26(4), pp. 1043–1059.

Kandi, V.R., Castronovo, F., Brittle, P., Mastrolembo Ventura, S. & Nikolic, D. (2020) ‘Assessing the Impact of a Construction Virtual Reality Game on Design Review Skills of Construction Students’, *Journal of Architectural Engineering*, 26(4), p. 04020035.

Kang, H. (2013) ‘The prevention and handling of the missing data’, *Korean Journal of Anesthesiology*, 64(5), p. 402.

Kang, J., Liu, M. & Qu, W. (2017) ‘Using gameplay data to examine learning behavior patterns in a serious game’, *Computers in Human Behavior*, 72pp. 757–770.

Kashive, N., Khanna, V.T., Kashive, K. & Barve, A. (2022) ‘Gamifying Employer Branding:

Attracting Critical Talent in Crisis Situations like COVID-19', *Journal of Promotion Management*, 28(4), pp. 487–514.

Kato, P. & de Klerk, S. (2017) 'Serious Games for Assessment: Welcome to the Jungle.', *Journal of Applied Testing Technology*, 18(S1), pp. 1–6.

Ke, F. (2016) 'Designing and integrating purposeful learning in game play: a systematic review', *Educational Technology Research and Development*, 64(2), pp. 219–244.

Kearney, K.G., Starkey, E.M. & Miller, S.R. (2022) 'Digitizing Product Dissection: A Case Study on Augmented Reality and Animation in Engineering Education', *Journal of Mechanical Design*, 144(8), .

Keaveney, S., Athanasopoulou, L., Siatras, V., Stavropoulos, P., Mourtzis, D. & Dowling, D.P. (2021) 'Development and Implementation of a Digital Manufacturing Demonstrator for Engineering Education', *Procedia CIRP*, 104pp. 1674–1679.

Kerawalla, L., Luckin, R., Seljeflot, S. & Woolard, A. (2006) '"Making it real": exploring the potential of augmented reality for teaching primary school science', *Virtual Reality*, 10(3–4), pp. 163–174.

Kerr, D. & Chung, G.K.W.K. (2012) 'Identifying Key Features of Student Performance in Educational Video Games and Simulations through Cluster Analysis', *JEDM - Journal of Educational Data Mining*, 4(1), pp. 144–182.

Khamis, H. (2008) 'Measures of Association: How to Choose?', *Journal of Diagnostic Medical Sonography*, 24(3), pp. 155–162.

Khukalenko, I.S., Kaplan-Rakowski, R., An, Y. & Iushina, V.D. (2022) 'Teachers' perceptions of using virtual reality technology in classrooms: A large-scale survey', *Education and Information Technologies*, pp. 1–23.

Kiili, K. (2005) 'Digital game-based learning: Towards an experiential gaming model', *Internet and Higher Education*, 8(1), pp. 13–24.

Kim, Y.J., Almond, R.G. & Shute, V.J. (2016) ‘Applying Evidence-Centered Design for the Development of Game-Based Assessments in Physics Playground’, *International Journal of Testing*, 16(2), pp. 142–163.

Kittur, J. & Islam, T. (2021) ‘Serious Games in Engineering: The Current State, Trends, and Future’, in *ASEE Annual Conference and Exposition*. [Online]. 2021

Kleinlogel, E.P., Renier, L., Schmid Mast, M. & Toma, C. (2020) ‘“I Think That I Made a Good Impression!”’, *Social Psychology*, 51(6), pp. 370–380.

de Klerk, S., Veldkamp, B.P. & Eggen, T.J.H.M. (2015) ‘Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example’, *Computers & Education*, 85pp. 23–34.

Klopfer, E., Osterweil, S. & Salen, K. (2009) *Moving learning games forward*. Cambridge MA: Educational Arcade.

Kock, N. & Hadaya, P. (2018) ‘Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods’, *Information Systems Journal*, 28(1), pp. 227–261.

Kolb, A.Y. & Kolb, D.A. (2005) ‘Learning Styles and Learning Spaces: Enhancing Experiential Learning in Higher Education’, *Academy of Management Learning & Education*, 4(2), pp. 193–212.

Krippendorff, K. (1989) ‘Content Analysis’, in E. Barnouw, G. Gerbner, W. Schramm, T. L. Worth, & L. Gross (eds.) , *International encyclopedia of communication*. [Online]. New York, NY: Oxford University Press. pp. 403–407.

Kumar, J.A. & Bervell, B. (2019) ‘Google Classroom for mobile learning in higher education: Modelling the initial perceptions of students’, *Education and Information Technologies*, 24(2), pp. 1793–1817.

Kumar, S., M, M., B, V., Singh, M. & K, C. (2019) ‘A brief survey of unsupervised agglomerative hierarchical clustering schemes’, *International Journal of Engineering &*

Technology, 8(1), pp. 29–37.

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*, 36pp. 143–153.

Landau, S., Leese, M., Stahl, D., Everitt, B.S., Landau, D.S. & Leese, D.M. (2011) *Cluster analysis*. Vol. 20. New Jersey: John Wiley.

Lave, J. & Wenger, E. (1991) *Situated Learning: Legitimate Peripheral Participation.*

Lee, E.A.-L. & Wong, K.W. (2014) 'Learning with desktop virtual reality: Low spatial ability learners are more positively affected', *Computers & Education,* 79pp. 49–58.

Lee, E.A.-L., Wong, K.W. & Fung, C.C. (2010) 'Learning with Virtual Reality: Its Effects on Students with Different Learning Styles', in Zhigeng Pan, Adrian David Cheok, Wolfgang Müller, Xiaopeng Zhang, & Kevin Wong (eds.) *Transactions on Edutainment IV*. [Online]. Heidelberg: Springer. pp. 79–90.

Lee, Y.-C.N., Shan, L.-T. & Chen, C.-H. (2013) 'System Development of Immersive Technology Theatre in Museum', in R. Shumaker (ed.) *Virtual, Augmented and Mixed Reality. Systems and Applications. VAMR 2013. Lecture Notes in Computer Science*. [Online]. Berlin: Springer.

Li, J., van der Spek, E.D., Feijs, L., Wang, F. & Hu, J. (2017) *Augmented Reality Games for Learning: A Literature Review,* in [Online]. pp. 612–626.

Limayem, Hirt & Cheung (2007) 'How Habit Limits the Predictive Power of Intention: The Case of Information Systems Continuance', *MIS Quarterly,* 31(4), p. 705.

Lin, Y. & Wang, S. (2019) 'The Study and Application of Adaptive Learning Method Based on Virtual Reality for Engineering Education', in Y Zhao, N. Barnes, B. Chen, R. Westermann, X. Kong, & C. Lin (eds.) *International Conference on Image and Graphics (Lecture Notes in Computer Science).* [Online]. Cham: Springer. pp. 372–383.

Liu, Y. & Wang, Z. (2014) ‘Positive Affect and Cognitive Control’, *Psychological Science*, 25(5), pp. 1116–1123.

Lock, J., Kim, B., Koh, K. & Wilcox, G. (2018) ‘Navigating the Tensions of Innovative Assessment and Pedagogy in Higher Education.’, *The Canadian Journal for the Scholarship of Teaching and Learning*, 9(1), .

Loh, C. & Sheng, Y. (2015) ‘Measuring the (dis-)similarity between expert and novice behaviors as serious games analytics’, *Education and Information Technologies*, 20(1), pp. 5–19.

Loh, C., 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*, 49pp. 147–155.

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*. [Online]. New York: Springer.

Loh, C.S. (2009) ‘Researching and Developing Serious Games as Interactive Learning Instructions’, *International Journal of Gaming and Computer-Mediated Simulations*, 1(4), pp. 1–19.

Loh, C.S. & Sheng, Y. (2013) ‘Performance metrics for serious games: Will the (real) expert please step forward?’, in *Proceedings of CGAMES’2013 USA*. [Online]. July 2013 IEEE. pp. 202–206.

Loh, C.S., Sheng, Y. & Ifenthaler, D. (2015) ‘Serious Games Analytics: Theoretical Framework’, in Christian Sebastian Loh, Yanyan Sheng, & Dirk Ifenthaler (eds.) *Serious Games Analytics*. [Online]. Cham: Springer International Publishing. pp. 3–29.

Lu, Y.-L. & Lien, C.-J. (2020) ‘Are They Learning or Playing? Students’ Perception Traits and Their Learning Self-Efficacy in a Game-Based Learning Environment’, *Journal of Educational Computing Research*, 57(8), pp. 1879–1909.

Luo, X. & Mojica Cabico, C.D. (2018) ‘Development and Evaluation of an Augmented Reality Learning Tool for Construction Engineering Education’, in *Construction Research Congress 2018*. [Online]. 29 March 2018 Reston, VA: American Society of Civil Engineers. pp. 149–159.

Mahmud, W., Hyder, O., Butt, J. & Aftab, A. (2011) ‘Dissection videos do not improve anatomy examination scores’, *Anatomical Sciences Education*, 4(1), pp. 16–21.

Malaquias, R.F.O., Malaquias, F.F.O. & Hwang, Y. (2018) ‘Understanding technology acceptance features in learning through a serious game’, *Computers in Human Behavior*, 87pp. 395–402.

Malheiro, B., Guedes, P., Silva, M.F. & Ferreira, P. (2019) ‘Fostering Professional Competencies in Engineering Undergraduates with EPS@ISEP’, *Education Sciences*, 9(2), p. 119.

Martín-Gutiérrez, J., Contero, M. & Alcañiz, M. (2015) ‘Augmented Reality to Training Spatial Skills’, *Procedia Computer Science*, 77pp. 33–39.

Martin-Gutierrez, J., Guinters, E. & Perez-Lopez, D. (2012) ‘Improving Strategy of Self-Learning in Engineering: Laboratories with Augmented Reality’, *Procedia - Social and Behavioral Sciences*, 51pp. 832–839.

Martín-Gutierrez, J., Trujillo, R.E.N. & Acosta-Gonzalez, M.M. (2013) ‘Augmented Reality Application Assistant for Spatial Ability Training. HMD vs Computer Screen Use Study’, *Procedia - Social and Behavioral Sciences*, 93pp. 49–53.

Martin, F. & Betrus, A.K.K. (2019) *Digital media for learning: Theories, processes, and solutions*. Cham: Springer International Publishing.

Martin, J.A., Regehr, G., Reznick, R., Macrae, H., Murnaghan, J., Hutchison, C. & Brown, M. (1997) ‘Objective structured assessment of technical skill (OSATS) for surgical residents’, *British Journal of Surgery*, 84(2), pp. 273–278.

Mavridis, A. & Tsiatsos, T. (2017) ‘Game-based assessment: investigating the impact on test anxiety and exam performance', *Journal of Computer Assisted Learning,* 33(2), pp. 137–150.

May, J.O. & Looney, S.W. (2020) 'Sample Size Charts for Spearman and Kendall Coefficients Methods for Sample Size Determination', *Biometrics & Biostatistics,* 11(1), pp. 5–11.

McArthur, J. (2022) 'Rethinking authentic assessment: work, well-being, and society', *Higher Education,* pp. 1–17.

McCord, K.H., Ayer, S.K., Perry, L.A., Patil, K.R., London, J.S., Khoury, V. & Wu, W. (2022) 'Student Approaches and Performance in Element Sequencing Tasks Using 2D and Augmented Reality Formats', *Education Sciences,* 12(4), p. 247.

McCusker, J., Almaghrabi, M. & Kucharski, B. (2018) 'Is a Virtual Reality-based Laboratory Experience a Viable Alternative to the Real Thing?', in *2018 ASEE Annual Conference & Exposition Proceedings*. [Online]. 2018 ASEE Conferences.

McMorran, C., Ragupathi, K. & Luo, S. (2017) 'Assessment and learning without grades? Motivations and concerns with implementing gradeless learning in higher education', *Assessment and Evaluation in Higher Education,* 42(3), pp. 361–377.

McPhee, I. & D'Esposito, M.E.W. (2018) 'Student assessment in higher education: embargo or empowerment?', *Journal of Applied Research in Higher Education,* 10(2), pp. 155–169.

McQuade, R., Ventura-Medina, E., Wiggins, S. & Anderson, T. (2020) 'Examining self-managed problem-based learning interactions in engineering education', *European Journal of Engineering Education,* 45(2), pp. 232–248.

Mehrfard, A., Fotouhi, J., Taylor, G., Forster, T., Navab, N. & Fuerst, B. (2019) 'A Comparative Analysis of Virtual Reality Head-Mounted Display Systems', *arXiv preprint,*

Menekse Dalveren, G.G. & Cagiltay, N.E. (2020) 'Distinguishing Intermediate and Novice

Surgeons by Eye Movements’, *Frontiers in Psychology*, 11(September), pp. 1–10.

Merrett, C. (2022) *Using case studies and build projects as authentic assessments in cornerstone courses*. Vol. 50.

Messick, S. (1987) ‘VALIDITY’. ETS Research Report Series

Miljanovic, M.A. & Bradbury, J.S. (2020) ‘GidgetML’, in *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering Education and Training*. [Online]. 27 June 2020 New York, NY, USA: ACM. pp. 184–192.

Mio, C., Ventura-Medina, E. & João, E. (2019) ‘Scenario-based eLearning to promote active learning in large cohorts: Students’ perspective’, *Computer Applications in Engineering Education*, 27(4), pp. 894–909.

Mirabolghasemi, M., Hamilton, M., Jones, P., Cole, D., Wall, E. & Jaradat, R. (2019) ‘Understanding the Effectiveness of Using VR to Support Teaching Drilling Trajectory Concepts’, in *2019 ASEE Annual Conference & Exposition Proceedings*. [Online]. 2019 ASEE Conferences.

Mislevy, R., Almond, R. & Lukas, J.F. (2003) *A Brief Introduction to Evidence-centered Design*.

Mislevy, R., Steinberg, L.S. & Almond, R.G. (2003) ‘Focus Article: On the Structure of Educational Assessments’, *Measurement: Interdisciplinary Research & Perspective*, 1(1), pp. 3–62.

Mislevy, R.J., Behrens, J.T., Dicerbo, K.E., Frezzo, D.C. & West, P. (2012) ‘Three Things Game Designers Need to Know About Assessment’, in *Assessment in Game-Based Learning*. [Online]. New York, NY: Springer New York. pp. 59–81.

Moseley, D., Baumfield, V., Elliott, J., Higgins, S., Miller, J., Newton, D.P. & Maggie Gregson (2005) *Frameworks for Thinking : A Handbook for Teaching and Learning*. Cambridge: Cambridge University Press.

Moskal, B.M., Leydens, J.A. & Pavelich, M.J. (2002) ‘Validity, Reliability and the Assessment of Engineering Education’, *Journal of Engineering Education*, 91(3), pp. 351–354.

Nelson, M. & Ahn, B. (2018) *Virtual Reality Activities for Teaching Engineering Students Professional Development Skills*,

Ninaus, M., Moeller, K., McMullen, J. & Kiili, K. (2017) ‘Acceptance of Game-Based Learning and Intrinsic Motivation as Predictors for Learning Success and Flow Experience’, *International Journal of Serious Games*, 4(3), .

Nitzl, C. (2016) ‘The use of partial least squares structural equation modelling (PLS-SEM) in management accounting research: Directions for future theory development’, *Journal of Accounting Literature*, 37(1), pp. 19–35.

Noraddin, E.M. & Kian, N.T. (2014) ‘Academics’ Attitudes Toward Using Digital Games for Learning & Teaching in Malaysia’, *Malaysian Online Journal of Educational Technology*, 2(4), pp. 1–21.

Norman, K.L. (2017) *Cyberpsychology: An Introduction to Human-Computer Interaction*. 2nd edition. Cambridge: Cambridge University Press.

Oren, M., Pedersen, S. & Butler-Purry, K.L. (2021) ‘Teaching Digital Circuit Design With a 3-D Video Game: The Impact of Using In-Game Tools on Students’ Performance’, *IEEE Transactions on Education*, 64(1), pp. 24–31.

Ostrander, J.K., Tucker, C.S., Simpson, T.W. & Meisel, N.A. (2018) ‘Evaluating the Effectiveness of Virtual Reality As an Interactive Educational Resource for Additive Manufacturing’, in *Volume 3: 20th International Conference on Advanced Vehicle Technologies; 15th International Conference on Design Education*. [Online]. 26 August 2018 American Society of Mechanical Engineers.

Ostrander, J.K., Tucker, C.S., Simpson, T.W. & Meisel, N.A. (2020) ‘Evaluating the Use of Virtual Reality to Teach Introductory Concepts of Additive Manufacturing’, *Journal of Mechanical Design*, 142(5), .

Pando-Garcia, J., Periañez-Cañadillas, I. & Charterina, J. (2016) 'Business simulation games with and without supervision: An analysis based on the TAM model', *Journal of Business Research,* 69(5), pp. 1731–1736.

Pando Cerra, P., Fernández Álvarez, H., Busto Parra, B. & Iglesias Cordera, P. (2022) 'Effects of Using Game-Based Learning to Improve the Academic Performance and Motivation in Engineering Studies', *Journal of Educational Computing Research,*

Park, B.J., Hunt, S.J., Martin, C., Nadolski, G.J., Wood, B.J. & Gade, T.P. (2020) 'Augmented and Mixed Reality: Technologies for Enhancing the Future of IR', *Journal of Vascular and Interventional Radiology,* 31(7), pp. 1074–1082.

Paro, M.R., Hersh, D.S. & Bulsara, K.R. (2022) 'History of Virtual Reality and Augmented Reality in Neurosurgical Training', *World Neurosurgery,* 167pp. 37–43.

Peeters, M.J. & Schmude, K.A. (2020) 'Learning Assessment vs Program Evaluation', *American Journal of Pharmaceutical Education,* 84(9), p. ajpe7938.

Pence, H.E. (2010) 'Smartphones, Smart Objects, and Augmented Reality', *The Reference Librarian,* 52(1–2), pp. 136–145.

Peng, X., Isaac, B. & Wilkins, R.T. (2012) 'Development of nanoscale Virtual Reality simulations for the teaching of nanotechnology', *ASEE Annual Conference and Exposition, Conference Proceedings,*

Perini, S., Luglietti, R., Margoudi, M., Oliveira, M. & Taisch, M. (2018a) 'Learning and motivational effects of digital game-based learning (DGBL) for manufacturing education –The Life Cycle Assessment (LCA) game', *Computers in Industry,* 102pp. 40–49.

Perini, S., Luglietti, R., Margoudi, M., Oliveira, M. & Taisch, M. (2018b) 'Learning and motivational effects of digital game-based learning (DGBL) for manufacturing education –The Life Cycle Assessment (LCA) game', *Computers in Industry,* 102pp. 40–49.

Perini, S., Oliveira, M., Margoudi, M. & Taisch, M. (2018) 'The Use of Digital Game Based Learning in Manufacturing Education – A Case Study', in P. Zaphiris & A. Ioannou (eds.) *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [Online]. Cham: Springer. pp. 185–199.

Plass, J.L., Homer, B.D. & Kinzer, C.K. (2015) 'Foundations of Game-Based Learning', *Educational Psychologist*, 50(4), pp. 258–283.

Prensky, M. (2003) 'Digital game-based learning', *Computers in Entertainment*, 1(1), p. 21.

Prensky, M. (2001) 'Digital Natives, Digital Immigrants Part 1', *On the Horizon*, 9(5), pp. 1–6.

Prince, M. (2004) 'Does Active Learning Work? A Review of the Research', *Journal of Engineering Education*, 93(3), pp. 223–231.

R. Sanchez, D., Rueda, A., Jimeno Jimènez, L. & Norouzi Nargesi, M. (2022) 'Commercial-off-the-Shelf (COTS) Games: Exploring the Applications of Games for Instruction and Assessment', in Branislav Sobota (ed.) *Computer Game Development*. [Online]. London: IntechOpen.

Radianti, J., Majchrzak, T.A., Fromm, J. & Wohlgenannt, I. (2020) 'A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda', *Computers & Education*, 147p. 103778.

Ramírez-Correa, P., Rondán-Cataluña, F.J., Arenas-Gaitán, J. & Martín-Velicia, F. (2019) 'Analysing the acceptation of online games in mobile devices: An application of UTAUT2', *Journal of Retailing and Consumer Services*, 50pp. 85–93.

Razak, A.A., Connolly, T. & Hainey, T. (2012) 'Teachers' Views on the Approach of Digital Games-Based Learning within the Curriculum for Excellence', *International Journal of Game-Based Learning*, 2(1), pp. 33–51.

Reeves, T.C. & Hedberg, J.G. (2009) 'Evaluation Strategies for Open and Distributed Learning Environments', in Christine Spratt & Paul Lajbcygier (eds.) *E-Learning Technologies and Evidence-Based Assessment Approaches*. [Online]. Hershey: IGI

Global. pp. 234–242.

Reuter, R., Hauser, F., Muckelbauer, D., Stark, T., Antoni, E., Mottok, J. & Wolff, C. (2019) ‘Using Augmented Reality in Software Engineering Education? First insights to a comparative study of 2D and AR UML modeling’, in *Proceedings of the Annual Hawaii International Conference on System Sciences*. [Online]. 2019 pp. 7798–7807.

Rice, J.W. (2007) ‘Assessing Higher Order Thinking in Video Games’, *Journal of Technology and Teacher Education*, 15(1), pp. 87–100.

Ritter, K.A., Borst, C.W. & Chambers, T.L. (2018) ‘Virtual solar energy center case studies’, *Computers in Education Journal*, 9(4), .

Roberts, D. (2019) ‘Higher education lectures: From passive to active learning via imagery?’, *Active Learning in Higher Education*, 20(1), pp. 63–77.

Roca González, C., García Domínguez, M., Martín-Gutierrez, J., Contero, M. & Alcañiz, M. (2012) ‘Training With Augmented Reality on Engineering Degrees’, in *ASME 2012 11th Biennial Conference on Engineering Systems Design and Analysis*. [Online]. 2 July 2012 American Society of Mechanical Engineers.

Rossado Espinoza, V.P., Cardenas-Salas, D., Cabrera, A. & Coronel, L. (2021) ‘Virtual Reality and BIM Methodology as Teaching- Learning Improvement Tools for Sanitary Engineering Courses’, *International Journal of Emerging Technologies in Learning (iJET)*, 16(06), p. 20.

Rourke, L. & Anderson, T. (2004) ‘Validity in quantitative content analysis’, *Educational Technology Research and Development*, 52pp. 5–18.

Rourke, L., Anderson, T., Garrison, D.R. & Archer, W. (2001) ‘Methodological Issues in the Content Analysis of Computer Conference Transcripts’, *International Journal of Artificial Intelligence in Education*, 12(July 2000), pp. 8–22.

Routledge, H. (2009) ‘Games-Based Learning in the Classroom and How it can Work!’, in Thomas Connolly, Mark Stansfield, & Liz Boyle (eds.) *Games-Based Learning*

Advancements for Multi-Sensory Human Computer Interfaces. [Online]. Pennsylvania: IGI Global. pp. 274–286.

Sagnier, C., Loup-Escande, E. & Vallery, G. (2020) ‘Effects of Gender and Prior Experience in Immersive User Experience with Virtual Reality’, in *Advances in Intelligent Systems and Computing*. [Online]. pp. 305–314.

Salah, B., Abidi, M., Mian, S., Krid, M., Alkhalefah, H. & Abdo, A. (2019) ‘Virtual Reality-Based Engineering Education to Enhance Manufacturing Sustainability in Industry 4.0’, *Sustainability*, 11(5), p. 1477.

Saldana, J. (2009) *Coding Manual for Qualitative Researchers*. SAGE Publications.

Saleh, N., Prakash, E. & Manton, R. (2014) ‘Measuring student acceptance of game based learning for game and technology education curriculum development’, in *2014 International Conference on Education Technologies and Computers (ICETC)*. [Online]. 25 September 2014 IEEE. pp. 79–85.

Samsudeen, S.N. & Mohamed, R. (2019) ‘University students’ intention to use e-learning systems’, *Interactive Technology and Smart Education*, 16(3), pp. 219–238.

Sánchez, A., Redondo, E. & Fonseca, D. (2012) ‘Developing an augmented reality application in the framework of architecture degree’, in *Proceedings of the 2012 ACM workshop on User experience in e-learning and augmented technologies in education - UXeLATE ’12*. [Online]. 2012 New York, New York, USA: ACM Press.

Seader, J.D., Henley, E.J. & Keith Roper, D. (2011) *Separation Process Principles: Chemical and Biochemical Operations*. 3rd edition. New Jersey: John Wiley & Sons, Inc.

Sevim-Cirak, N. & Yıldırım, Z. (2020) ‘Educational use and motivational elements of simulation games for mining engineering students: a phenomenological study’, *European Journal of Engineering Education*, 45(4), pp. 550–564.

Sherman, W.R. & Craig, A.B. (2003) *Understanding Virtual Reality:* Amsterdam: Elsevier.

Shernoff, D.J., Ryu, J.-C., Ruzek, E., Coller, B. & Prantil, V. (2020) ‘The Transportability of a Game-Based Learning Approach to Undergraduate Mechanical Engineering Education: Effects on Student Conceptual Understanding, Engagement, and Experience’, *Sustainability*, 12(17), p. 6986.

Shirazi, A. & Behzadan, A.H. (2015) ‘Content delivery using augmented reality to enhance students’ performance in a building design and assembly project’, *Advances in Engineering Education*, 4(3), .

Shiue, Y.-M., Hsu, Y.-C. & Liang, Y.-C. (2017) ‘Modeling the continuance usage intention of game-based learning in the context of collaborative Learning’, in *IEEE International Conference on Applied System Innovation*. [Online]. 24 September 2017 IEEE. pp. 1106–1109.

Shute, Rahimi, S. & Emihovich, B. (2017) ‘Assessment for Learning in Immersive Environments’, in D. Liu, C. Dede, R. Huang, & J. Richards (eds.) *Virtual, Augmented, and Mixed Realities in Education.* [Online]. Singapore: Springer. pp. 71–87.

Shute, V., Ke, F. & Wang, L. (2017) ‘Assessment and Adaptation in Games’, in Pieter Wouters & Herre van Oostendorp (eds.) *Instructional Techniques to Facilitate Learning and Motivation of Serious Games*. [Online]. Cham: Springer International Publishing. pp. 59–78.

Shute, V. & Ventura, M. (2013) *Stealth Assessment: Measuring and Supporting Learning in Video Games*. London: MIT Press.

Shute, V.J. & Ke, F. (2012) ‘Games, Learning, and Assessment’, in *Assessment in Game-Based Learning*. [Online]. New York, NY: Springer New York. pp. 43–58.

Shute, V.J. & Rahimi, S. (2021) ‘Stealth assessment of creativity in a physics video game’, *Computers in Human Behavior*, 116p. 106647.

Shute, V.J., Rahimi, S. & Sun, C. (2018) *Measuring & Supporting Learning in Educational Games*, pp. 192–210.

Shute, V.J., Ventura, M., Bauer, M. & Zapata-Rivera (2009) ‘Melding the Power of Serious Games and Embedded Assessment to Monitor and Foster Learning: Flow and Grow’, in U. Ritterfield, M. Cody, & P. Vorderer (eds.) *Serious Games: Mechanisms and Effects*. [Online]. New York: Routledge. pp. 296–321.

Shute, V.J., Ventura, M. & Ke, F. (2015) ‘The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills’, *Computers & Education,* 80pp. 58–67.

Siddique, Z., Ling, C., Roberson, P., Xu, Y. & Geng, X. (2013) ‘Facilitating Higher-Order Learning Through Computer Games’, *Journal of Mechanical Design*, 135(12), .

Singh, G., Mantri, A., Sharma, O., Dutta, R. & Kaur, R. (2019) ‘Evaluating the impact of the augmented reality learning environment on electronics laboratory skills of engineering students’, *Computer Applications in Engineering Education,* 27(6), pp. 1361–1375.

Slimani, A., Elouaai, F., Elaachak, L., Bakkali Yedri, O., Bouhorma, M. & Sbert, M. (2018) ‘Learning Analytics Through Serious Games: Data Mining Algorithms for Performance Measurement and Improvement Purposes’, *International Journal of Emerging Technologies in Learning (iJET),* 13(01), p. 46.

Smith, G., Shute, V. & Muenzenberger, A. (2019) ‘Designing and Validating a Stealth Assessment for Calculus Competencies.’, *Journal of Applied Testing Technology,* 20(S1), pp. 52–59.

Smith, S. & Chan, S. (2017) ‘Collaborative and Competitive Video Games for Teaching Computing in Higher Education’, *Journal of Science Education and Technology,* 26(4), pp. 438–457.

Solmaz, S., Dominguez Alfaro, J.L., Santos, P., Van Puyvelde, P. & Van Gerven, T. (2021) ‘A practical development of engineering simulation-assisted educational AR environments’, *Education for Chemical Engineers,* 35pp. 81–93.

Solmaz, S. & Van Gerven, T. (2021) ‘Automated integration of extract-based CFD results with AR/VR in engineering education for practitioners’, *Multimedia Tools and Applications,* Squire, K. (2006) ‘From content to context: Videogames as designed experience’, *Educational Researcher*, 35(8), pp. 19–29.

Squire, K. (2003) ‘Video Games in Education’, *International Journal of Intelligent Games & Simulation*, 2(1), .

Stieler-Hunt, C. & Jones, C.M. (2015) ‘Educators who believe: understanding the enthusiasm of teachers who use digital games in the classroom’, *Research in Learning Technology*, 23(September), .

Suescún, E., Camapio do Prado Leite, J.C., Werneck, V., Mazo, R., Vallejo, P., Toro, M., Velasquez, D., Cardona, J.F. & Rincón, R. (2018) ‘SimulES-W: A Collaborative Game to Improve Software Engineering Teaching’, *Computación y Sistemas*, 22(3), .

Sun, C.-T., Chou, K.-T. & Yu, H.C. (2022) ‘Relationship between digital game experience and problem-solving performance according to a PISA framework’, *Computers & Education*, 186p. 104534.

Sutherland, I.E. (1965) ‘The Ultimate Display’, in *IFIPS Congress*. [Online]. 1965 New York: . pp. 506–508.

Suzuki, K., Shibuya, T. & Kanagawa, T. (2021) ‘Effectiveness of a game-based class for interdisciplinary energy systems education in engineering courses’, *Sustainability Science*, 16(2), pp. 523–539.

Svinicki, M.D. (2004) ‘Authentic assessment: Testing in reality’, *New Directions for Teaching and Learning*, (100), pp. 23–29.

Sweet, S.J. & Rupp, A.A. (2012) ‘Using the ECD Framework to Support Evidentiary Reasoning in the Context of a Simulation Study for Detecting Learner Differences in Epistemic Games’, *Journal of Educational Data Mining*, 4(1), pp. 183–223.

Thanasi-Boçe, M. (2020) ‘Enhancing students’ entrepreneurial capacity through marketing simulation games’, *Education + Training*, 62(9), pp. 999–1013.

Tixier, A. & Albert, A. (2013) ‘Teaching construction hazard recognition through high fidelity augmented reality’, in *2013 ASEE Annual Conference & Exposition Proceedings*. [Online]. 2013 ASEE Conferences. pp. 23.1139.1-23.1139.15.

Topalli, D. & Cagiltay, N.E. (2019) ‘Classification of Intermediate and Novice Surgeons’ Skill Assessment Through Performance Metrics’, *Surgical Innovation*, 26(5), pp. 621–629.

Toyoda, R., Abegão, F.R., Gill, S. & Glassey, J. (2020) ‘Perceptions of Professional Chemical Engineers toward Immersive Virtual Reality in Health and Safety Training’, in *The European Conference on Education 2020: Official Conference Proceedings*. [Online]. 2020 pp. 273–284.

Tsai, M.-J., Huang, L.-J., Hou, H.-T., Hsu, C.-Y. & Chiou, G.-L. (2016) ‘Visual behavior, flow and achievement in game-based learning’, *Computers & Education*, 98pp. 115–129.

Turkan, Y., Radkowski, R., Karabulut-Ilgu, A., Behzadan, A.H. & Chen, A. (2017) ‘Mobile augmented reality for teaching structural analysis’, *Advanced Engineering Informatics*, 34(July), pp. 90–100.

Tzima, S., Styliaras, G. & Bassounas, A. (2019) ‘Augmented Reality Applications in Education: Teachers Point of View’, *Education Sciences*, 9(2), p. 99.

Udeozor, C., Russo Abegão, F. & Glassey, J. (2022) ‘An Evaluation of the Relationship Between Perceptions and Performance of Students in a Serious Game’, *Journal of Educational Computing Research*, 60(2), pp. 322–351.

Udeozor, C., Russo Abegão, F. & Glassey, J. (2021) ‘Exploring Log Data for Behaviour and Solution Pattern Analyses in a Serious Game’, in Ugur Bakan & S Berkeley (eds.) *Gamification and Social Networks in Education*. [Online]. MacroWorld Pub. Ltd.

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*, pp. 1–19.

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), pp. 175–194.

Unglert, K., Radić, V. & Jellinek, A.M. (2016) ‘Principal component analysis vs. self-organizing maps combined with hierarchical clustering for pattern recognition in volcano seismic spectra’, *Journal of Volcanology and Geothermal Research*, 320pp. 58–74.

Urbano, D., Fátima Chouzal, M. & Restivo, M.T. (2020) ‘Evaluating an online augmented reality puzzle for DC circuits: Students’ feedback and conceptual knowledge gain’, *Computer Applications in Engineering Education*, 28(5), pp. 1355–1368.

Urbina Coronado, P.D., Demeneghi, J.A.A., Ahuett-Garza, H., Orta Castañon, P. & Martínez, M.M. (2022) ‘Representation of machines and mechanisms in augmented reality for educative use’, *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 16(2), pp. 643–656.

Vásquez-Carbonell, M. (2022a) ‘A Systematic Literature Review of Augmented Reality in Engineering Education : Hardware, Software, Student Motivation & Development Recommendations.’, *Digital Education Review*, (41), pp. 249–267.

Vásquez-Carbonell, M. (2022b) ‘A Systematic Literature Review of Virtual Reality in Engineering Education’, *International Journal of Virtual and Personal Learning Environments*, 12(1), pp. 1–18.

Veide, Z. & Strozheva, V. (2019) ‘Triggering the Students’ Positive Attitude for the Studies of Engineering Graphics Courses through the Augmented Reality Content’, in *ENVIRONMENT. TECHNOLOGIES. RESOURCES. Proceedings of the International Scientific and Practical Conference*. [Online]. 20 June 2019 Rezekne Higher Education Institution. pp. 242–246.

Venkatesh, Thong & Xu (2012) ‘Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology’, *MIS Quarterly*, 36(1), p. 157.

Venkatesh, V., Morris, M.G., Davis, G.B. & Davis, F.D. (2003) ‘User acceptance of information technology: Toward a unified view’, *MIS Quarterly: Management Information Systems*, 27(3), pp. 425–478.

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), pp. 328–376.

Vert, S. & Andone, D. (2019) ‘Virtual Reality Authoring Tools for Educators’, *ITM Web of Conferences*, 29(2), p. 03008.

Villarroel, V., Boud, D., Bloxham, S., Bruna, D. & Bruna, C. (2019) ‘Using principles of authentic assessment to redesign written examinations and tests’, *Innovations in Education and Teaching International*, 57(1), pp. 1–12.

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

Wallner, G. & Kriglstein, S. (2012) ‘A spatiotemporal visualization approach for the analysis of gameplay data’, in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI ’12*. [Online]. 2012 New York, New York, USA: ACM Press. p. 1115.

Wang, F. & Hannafin, M.J. (2005) ‘Design-based research and technology-enhanced learning environments’, *Educational Technology Research and Development*, 53(4), .

Wang, H.Y. & Wang, Y.S. (2008) ‘Gender differences in the perception and acceptance of online games’, *British Journal of Educational Technology*, 39(5), pp. 787–806.

Wang, Y.-Y., Wang, Y.-S. & Jian, S.-E. (2020) ‘Investigating the Determinants of Students’ Intention to Use Business Simulation Games’, *Journal of Educational Computing Research*, 58(2), pp. 433–458.

Westera, W. (2019) ‘Why and how serious games can become far more effective: Accommodating productive learning experiences, learner motivation and the monitoring of learning gains', *Educational Technology and Society*, 22(1), pp. 59–69.

Westera, W., Nadolski, R. & Hummel, H. (2014) 'Serious Gaming Analytics: What Students´ Log Files Tell Us about Gaming and Learning', *International Journal of Serious Games*, 1(2), .

Westera, W., Prada, R., Mascarenhas, S., Santos, P.A., Dias, J., Guimarães, M., Georgiadis, K., Nyamsuren, E., Bahreini, K., Yumak, Z., Christyowidiasmoro, C., Dascalu, M., Gutu-Robu, G. & Ruseti, S. (2020) 'Artificial intelligence moving serious gaming: Presenting reusable game AI components', *Education and Information Technologies*, 25(1), pp. 351–380.

Whitton, N. (2009) 'Learning and Teaching with Computer Games in Higher Education', in Thomas Connolly, Mark Stansfield, & Boyle Liz (eds.) *Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces*. [Online]. IGI Global. pp. 18–33.

Wiggins, G. (2012) 'Seven Keys to Effective Feedback', *Feedback*, 70(1), pp. 10–16.

Wolf, M.J.P. (2001) *Genre and the Video Game*. Austin: University of Texas Press.

Wouters, P. & van der Meulen, E.S. (2020) 'The Role of Learning Styles in Game-Based Learning', *International Journal of Game-Based Learning*, 10(1), pp. 54–69.

Wouters, P. & van Oostendorp, H. (2017) 'Overview of Instructional Techniques to Facilitate Learning and Motivation of Serious Games', in Pieter Wouters & Herre van Oostendorp (eds.) *Instructional Techniques to Facilitate Learning and Motivation of Serious Games*. [Online]. Cham: Springer. pp. 1–16.

Wu, H.-K., Lee, S.W.-Y., Chang, H.-Y. & Liang, J.-C. (2013) 'Current status, opportunities and challenges of augmented reality in education', *Computers & Education*, 62pp. 41–49.

Wu, W.-H., Hsiao, H.-C., Wu, P.-L., Lin, C.-H. & Huang, S.-H. (2012) 'Investigating the learning-theory foundations of game-based learning: a meta-analysis', *Journal of*

Computer Assisted Learning, 28(3), pp. 265–279.

Wyatt-Smith, C. & Cumming, J.J. (2009) 'Fairness in Assessment', in Claire Wyatt-Smith & J. Joy Cumming (eds.) *Educational Assessment in the 21st Century: Connecting Theory and Practice*. [Online]. Dordrecht: Springer Netherlands.

Xenos, M. & Velli, V. (2020) 'A Serious Game for Introducing Software Engineering Ethics to University Students', in *International Conference on Interactive Collaborative Learning*. [Online]. 2020 pp. 579–588.

Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F.D., Li, C., Li, W., Huang, H., Akdere, M., Mousas, C. & Yu, L.-F. (2021) 'A Review on Virtual Reality Skill Training Applications', *Frontiers in Virtual Reality*, 2(April), pp. 1–19.

Yang, H., Yang, S. & Isen, A.M. (2013) 'Positive affect improves working memory: Implications for controlled cognitive processing', *Cognition & Emotion*, 27(3), pp. 474–482.

Yang, J. & Quadir, B. (2018) 'Effects of prior knowledge on learning performance and anxiety in an English learning online role-playing game', *Educational Technology and Society*, 21(3), pp. 174–185.

Youngblut, C. (1998) *Educational Uses of Virtual Reality Technology.*

Yue, W.S. & Tze, H.Y. (2015) 'Effectiveness of MMORPGS in Enhancing Student Interaction', in *2015 IEEE 15th International Conference on Advanced Learning Technologies*. [Online]. 14 July 2015 IEEE. pp. 152–154.

Zhao, D., Muntean, C.H., Chis, A.E. & Muntean, G.-M. (2021) 'Learner Attitude, Educational Background, and Gender Influence on Knowledge Gain in a Serious Games-Enhanced Programming Course', *IEEE Transactions on Education*, 64(3), pp. 308–316.

Zia, A. & Essa, I. (2018) 'Automated surgical skill assessment in RMIS training', *International Journal of Computer Assisted Radiology and Surgery*, 13(5), pp. 731–739.

Zimmerman, B.J. & Schunk, Dale H. (2011) ‘Self-Regulated Learning and Performance: An Introduction and an Overview’, in D. H. Schunk & Barry J. Zimmerman (eds.) *Handbook of Self-Regulation of Learning*. [Online]. New York: Routledge. p. 530.

Zulfiqar, S., Al-reshidi, H.A., Al Moteri, M.A., Feroz, H.M.B., Yahya, N. & Al-Rahmi, W.M. (2021) ‘Understanding and Predicting Students’ Entrepreneurial Intention through Business Simulation Games: A Perspective of COVID-19’, *Sustainability*, 13(4), p. 1838.

Appendices

Appendix 2.A: Mediums of Publication for the DG papers

Journals	Conferences	Books
Computers in Industry	IEEE Frontiers in Education Conference	Lecture Notes in Computer Science
Journal of Computing in Civil Engineering	International Conference on Engineering, Technology and Innovation (ICE)	
Sustainability	IEEE Transactions on Systems, Man, and Cybernetics: Systems	
Journal of Science Education and Technology	IEEE Global Engineering Education Conference (EDUCON)	
Information	International Conference on Software Engineering	
Journal of Mechanical Design	International Conference on Information Technology Based Higher Education and Training	
IEEE Transactions on Education	International Conference on Interactive Collaborative Learning	
Computación y Sistemas		
Computers in Education		
Journal of Educational Computing Research		

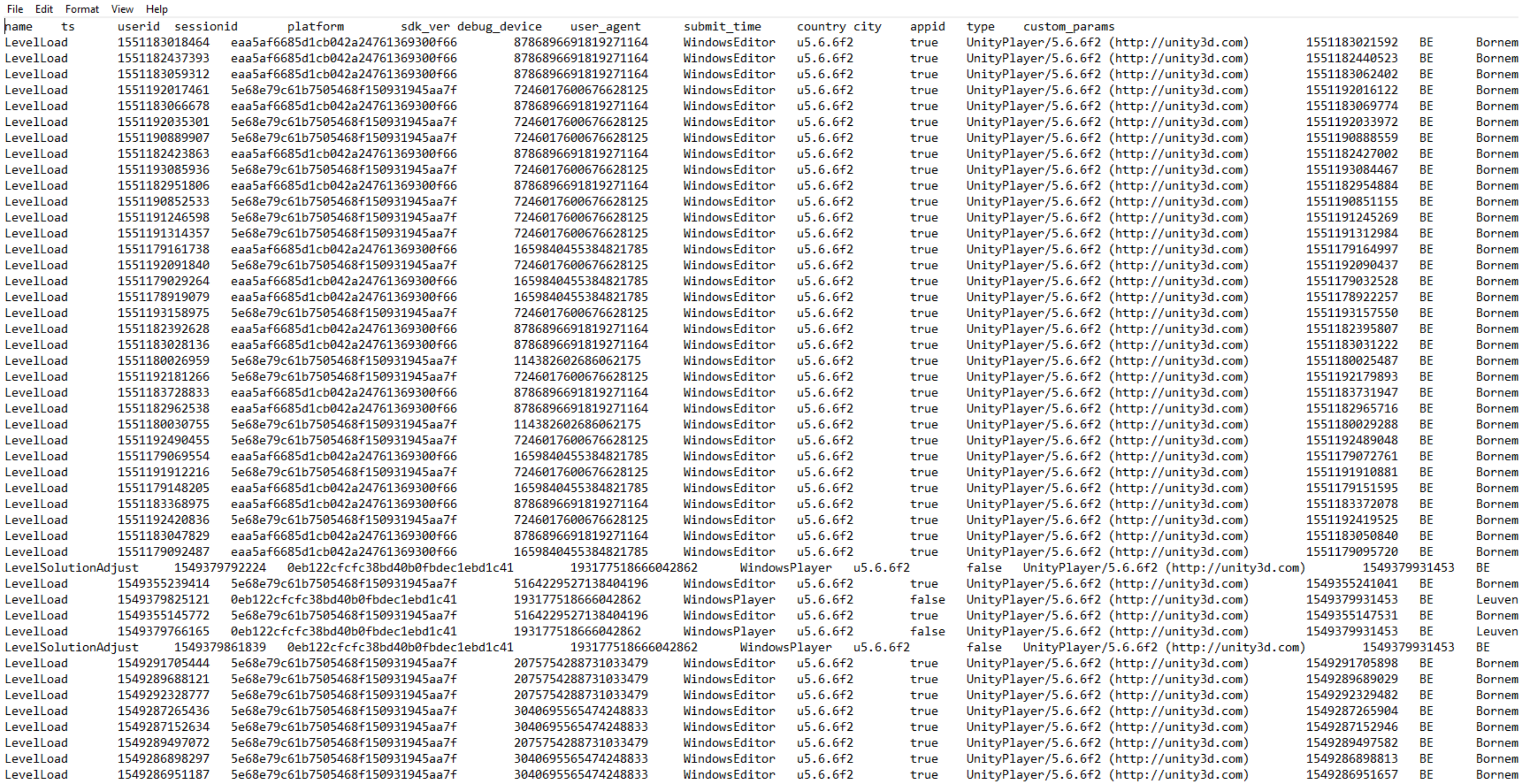
Appendix 2.B: Mediums of Publication for the VR papers

Journals	**Conferences**	**Books**
Computer Applications in Engineering Education	ASEE Annual Conference and Exposition Proceedings	Lecture Notes in Computer Science
Computers in Education Journal	International Conference on Computer Supported Education	
Sustainability	International Conference on Advanced Vehicle Technologies; International Conference on Design Education	
Alexandria Engineering Journal		
Journal of Architectural Engineering		
Journal of Mechanical Design		
Multimedia Tools and Applications		
Journal of Surveying Engineering		
International Journal of Emerging Technologies in Learning (iJET)		
Technology, Knowledge and Learning		
Engineering, Construction and Architectural Management		
Virtual Reality		

Appendix 2.C: Mediums of Publication for the AR papers

Journals	**Conferences**	**Books**
Computer Applications in Engineering Education	ASEE Annual Conference and Exposition Proceedings	Lecture Notes in Computer Science
Computers in Education Journal	International Conference on Computer Supported Education	
Alexandria Engineering Journal	International Conference on Advanced Vehicle Technologies; International Conference on Design Education	
Sustainability		
Journal of Architectural Engineering		
Journal of Mechanical Design		
Multimedia Tools and Applications		
Journal of Surveying Engineering		
International Journal of Emerging Technologies in Learning (iJET)		
Technology, Knowledge and Learning		
Engineering, Construction and Architectural Management		
Virtual Reality		

Appendix 4.A: Sample of Log Data from CosmiClean Gameplay



```
name    ts      userid  sessionid       platform        sdk_ver debug_device    user_agent      submit_time     country city    appid   type    custom_params
LevelLoad       1551183018464   eaa5af6685d1cb042a24761369300f66        8786896691819271164     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551183021592   BE      Bornem
LevelLoad       1551182437393   eaa5af6685d1cb042a24761369300f66        8786896691819271164     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551182440523   BE      Bornem
LevelLoad       1551183059312   eaa5af6685d1cb042a24761369300f66        8786896691819271164     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551183062402   BE      Bornem
LevelLoad       1551192017461   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551192016122   BE      Bornem
LevelLoad       1551183066678   eaa5af6685d1cb042a24761369300f66        8786896691819271164     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551183069774   BE      Bornem
LevelLoad       1551192035301   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551192033972   BE      Bornem
LevelLoad       1551190889907   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551190888559   BE      Bornem
LevelLoad       1551182423863   eaa5af6685d1cb042a24761369300f66        8786896691819271164     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551182427002   BE      Bornem
LevelLoad       1551193085936   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551193084467   BE      Bornem
LevelLoad       1551182951806   eaa5af6685d1cb042a24761369300f66        8786896691819271164     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551182954884   BE      Bornem
LevelLoad       1551190852533   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551190851155   BE      Bornem
LevelLoad       1551191246598   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551191245269   BE      Bornem
LevelLoad       1551191314357   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551191312984   BE      Bornem
LevelLoad       1551179161738   eaa5af6685d1cb042a24761369300f66        1659840455384821785     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551179164997   BE      Bornem
LevelLoad       1551192091840   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551192090437   BE      Bornem
LevelLoad       1551179029264   eaa5af6685d1cb042a24761369300f66        1659840455384821785     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551179032528   BE      Bornem
LevelLoad       1551178919079   eaa5af6685d1cb042a24761369300f66        1659840455384821785     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551178922257   BE      Bornem
LevelLoad       1551193158975   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551193157550   BE      Bornem
LevelLoad       1551182392628   eaa5af6685d1cb042a24761369300f66        8786896691819271164     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551182395807   BE      Bornem
LevelLoad       1551183028136   eaa5af6685d1cb042a24761369300f66        8786896691819271164     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551183031222   BE      Bornem
LevelLoad       1551180026959   5e68e79c61b7505468f150931945aa7f        114382602686062175      WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551180025487   BE      Bornem
LevelLoad       1551192181266   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551192179893   BE      Bornem
LevelLoad       1551183728833   eaa5af6685d1cb042a24761369300f66        8786896691819271164     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551183731947   BE      Bornem
LevelLoad       1551182962538   eaa5af6685d1cb042a24761369300f66        8786896691819271164     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551182965716   BE      Bornem
LevelLoad       1551180030755   5e68e79c61b7505468f150931945aa7f        114382602686062175      WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551180029288   BE      Bornem
LevelLoad       1551192490455   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551192489048   BE      Bornem
LevelLoad       1551179069554   eaa5af6685d1cb042a24761369300f66        1659840455384821785     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551179072761   BE      Bornem
LevelLoad       1551191912216   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551191910881   BE      Bornem
LevelLoad       1551179148205   eaa5af6685d1cb042a24761369300f66        1659840455384821785     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551179151595   BE      Bornem
LevelLoad       1551183368975   eaa5af6685d1cb042a24761369300f66        8786896691819271164     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551183372078   BE      Bornem
LevelLoad       1551192420836   5e68e79c61b7505468f150931945aa7f        7246017600676628125     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551192419525   BE      Bornem
LevelLoad       1551183047829   eaa5af6685d1cb042a24761369300f66        8786896691819271164     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551183050840   BE      Bornem
LevelLoad       1551179092487   eaa5af6685d1cb042a24761369300f66        1659840455384821785     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1551179095720   BE      Bornem
LevelSolutionAdjust     1549379792224   0eb122cfcfc38bd40b0fbdec1ebd1c41        193177518666042862      WindowsPlayer   u5.6.6f2        false   UnityPlayer/5.6.6f2 (http://unity3d.com)        1549379931453   BE
LevelLoad       1549355239414   5e68e79c61b7505468f150931945aa7f        5164229527138404196     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1549355241041   BE      Bornem
LevelLoad       1549379825121   0eb122cfcfc38bd40b0fbdec1ebd1c41        193177518666042862      WindowsPlayer   u5.6.6f2        false   UnityPlayer/5.6.6f2 (http://unity3d.com)        1549379931453   BE      Leuven
LevelLoad       1549355145772   5e68e79c61b7505468f150931945aa7f        5164229527138404196     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1549355147531   BE      Bornem
LevelLoad       1549379766165   0eb122cfcfc38bd40b0fbdec1ebd1c41        193177518666042862      WindowsPlayer   u5.6.6f2        false   UnityPlayer/5.6.6f2 (http://unity3d.com)        1549379931453   BE      Leuven
LevelSolutionAdjust     1549379861839   0eb122cfcfc38bd40b0fbdec1ebd1c41        193177518666042862      WindowsPlayer   u5.6.6f2        false   UnityPlayer/5.6.6f2 (http://unity3d.com)        1549379931453   BE
LevelLoad       1549291705444   5e68e79c61b7505468f150931945aa7f        2075754288731033479     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1549291705898   BE      Bornem
LevelLoad       1549289688121   5e68e79c61b7505468f150931945aa7f        2075754288731033479     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1549289689029   BE      Bornem
LevelLoad       1549292328777   5e68e79c61b7505468f150931945aa7f        2075754288731033479     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1549292329482   BE      Bornem
LevelLoad       1549287265436   5e68e79c61b7505468f150931945aa7f        3040695565474248833     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1549287265904   BE      Bornem
LevelLoad       1549287152634   5e68e79c61b7505468f150931945aa7f        3040695565474248833     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1549287152946   BE      Bornem
LevelLoad       1549289497072   5e68e79c61b7505468f150931945aa7f        2075754288731033479     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1549289497582   BE      Bornem
LevelLoad       1549286898297   5e68e79c61b7505468f150931945aa7f        3040695565474248833     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1549286898813   BE      Bornem
LevelLoad       1549286951187   5e68e79c61b7505468f150931945aa7f        3040695565474248833     WindowsEditor   u5.6.6f2        true    UnityPlayer/5.6.6f2 (http://unity3d.com)        1549286951657   BE      Bornem
```

Appendix 7.A: MCQs for the VR LaboSafe game (from Chan *et al.*, 2021)

Risks in the laboratory

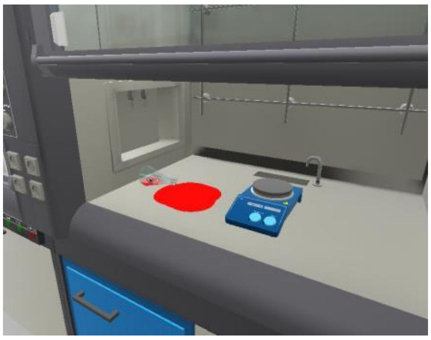
Risk 1: Overcrowded fume hood

Location	Fume hoods
Items	Bidon waste material CMR, bidon chloroform
Which hazard types are present in this photo?	1. Health
	2. Physical
	3. Environmental
	4. Failure of preventive/protective equipment
Correct hazards	1, 4
What are the possible consequences of this risk?	a. Risk of corrosive burn injury
	b. Insufficient ventilation of the fume hood
	c. Will cause inhalation problems
	d. Risk of explosion
Correct consequences	b, c
Hint	Sometimes it is easier to store everything in one fume hood ...
Safety information	Keep the fume hoods clean and unobstructed to avoid blocking the ventilation stream
Correct procedure	

Risk 2: Food drinks in laboratory

Location	Work bench tops
Items	Mug of coffee
Which hazard types are present in this photo?	1. Health
	2. Physical
	3. Environmental
	4. Failure of preventive/protective equipment
Correct hazards	1
What are the possible consequences of this risk?	a. Exposure of toxic fumes
	b. Can cause fire
	c. Risk of chemical contamination of food
	d. Burn injury by high temperature
Correct consequences	c
Hint	I smell coffee in the lab
Safety information	Do not eat, drink or store food products in the laboratory
Correct procedure	Throw away coffee

Risk 3: Flammable spill near heating plate



Location	Fume hoods
Items	Spill of isopropyl ether, electrical heating plate
Which hazard types are present in this photo?	1. Health
	2. Physical
	3. Environmental
	4. Failure of preventive/protective equipment
Correct hazards	2
What are the possible consequences of this risk?	a. Exposure of toxic fumes
	b. Corrosive burn injury on the skin
	c. Chance of ignition causing fire
	d. Environmental pollution
Correct consequences	c
Hint	Heat or electrical sparks and isopropyl ether do not go well together
Safety information	Keep flammable chemicals away from any ignition sources (e.g., heating plates, Bunsen burner, etc.)
Correct procedure	Clean up spill

Risk 4: Boxes in front of emergency shower

Location	Emergency shower
Items	Cardboard boxes
Which hazard types are present in this photo?	1. Physical hazard
	2. Ergonomic hazard
	3. Failure of emergency measures
	4. Environmental hazard
Correct hazards	3
What are the possible consequences of this risk?	a. Unable to easily access the emergency shower
	b. The boxes will get wet
	c. It makes the laboratory difficult to clean
	d. The boxes will catch on fire
Correct consequences	a
Hint	No one uses the shower. It won't be a problem to place something in front
Safety information	Keep emergency safety equipment unobstructed so they are easily reachable when there is an emergency
Correct procedure	Move boxes to the storage room without obstruction

Risk 5: Boxes in front of emergency exit

Location	Emergency exit
Items	Cardboard boxes
Which hazard types are present in this photo?	1. Physical hazard
	2. Ergonomic hazard
	3. Failure of emergency measures
	4. Environmental hazard
Correct hazards	3
What are the possible consequences of this risk?	a. People cannot enter the laboratory
	b. Finding the boxes is more difficult in this corner
	c. This corner cannot be easily cleaned
	d. Emergency exit is obstructed in case of evacuation
Correct consequences	d
Hint	There was so much delivery of packages this week. There was no time to store the package boxes anywhere else
Safety information	Keep evacuation routes unobstructed so that evacuation goes as fast as possible in case of an emergency
Correct procedure	Move boxes to the storage room without obstruction

Risk 6: Flammable product on the floor

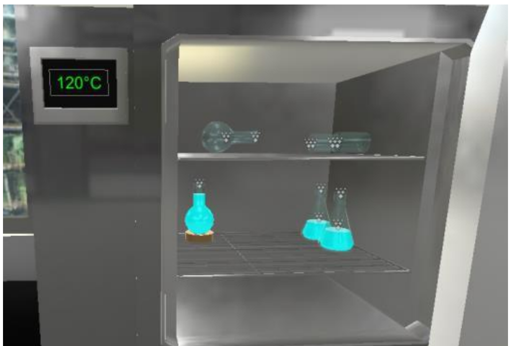
Location	Floors
Items	Bidon acetone
Which hazard types are present in this photo?	1. Physical
	2. Health
	3. Environmental
	4. Ergonomic hazard
Correct hazards	1, 4
What are the possible consequences of this risk?	a. Toxic fumes causes inhalation problems
	b. Can be easily knocked over, potentially causing fire
	c. High possibility of explosion
	d. Difficult to clean the floor
Correct consequences	b
Hint	To avoid risk that chemical containers fall from a high place, these are sometimes placed on the floor
Safety information	Do not place any chemical product on the floor. These can be easily knocked over by so
Correct procedure	Move to flammable cabine

Risk 7: Food in the laboratory fridge

Location	Normal or explosion-proof fridge
Items	Food container with spaghetti Bolognese lunch
Which hazard types are present in this photo?	1. Physical
	2. Health
	3. Environmental
	4. Failure of preventive/protective equipment
Correct hazards	2
What are the possible consequences of this risk?	a. Insufficient cooling of the fridge, causing products to heat up and explode
	b. Food will get spoiled, making it unsafe to eat
	c. Someone will steal this lunchbox
	d. Risk of chemical contamination of food
Correct consequences	d
Hint	Spaghetti leftovers cannot be stored at room temperature
Safety information	Do not eat, drink or store food products in the laboratory
Correct procedure	Throw away food in the trash bin

Risk 8: Flammable liquid in the non-explosion-proof fridge

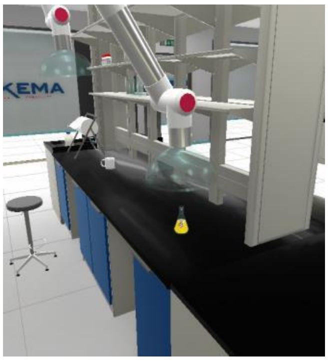
Location	Normal fridge
Items	Bottle of diethyl ether
Which hazard types are present in this photo?	1. Physical
	2. Health
	3. Environmental
	4. Failure of preventive/protective equipment
Correct hazards	1, 4
What are the possible consequences of this risk?	a. Someone will grab the wrong chemical product by mistake
	b. Fridge can malfunction and ignite the chemical causing fire or explosion
	c. Chemical can solidify and expand, causing the glass to break
	d. Fridge can malfunction and make the chemical ineffective
Correct consequences	b
Hint	Some laboratory fridges can generate electrical sparks inside
Safety information	Flammable chemicals must be stored in explosion-proof fridges to prevent possible ignitions
Correct procedure	Move to explosion-proof fridge



Risk 9: Cork flask support in oven

Location	Oven
Items	Cork flask support with round flask on top
Which hazard types are present in this photo?	1. Physical
	2. Health
	3. Environmental
	4. Failure of preventive/protective equipment
Correct hazards	1
What are the possible consequences of this risk?	a. Cork flask support will catch on fire in the oven
	b. Glassware can break at high temperature
	c. Evaporation of toxic fumes causes inhalation problems
	d. Heavy smoke will appear
Correct consequences	a, d
Hint	I smell something burned. You should look in the oven
Safety information	Avoid combustible materials in laboratory ovens
Correct procedure	Move outside the oven

Risk 10: Highly toxic chemical under local ventilation



Location	Local ventilation on workbench top
Items	Erlenmeyer flask of acrylonitrile
Which hazard types are present in this photo?	1. Physical
	2. Health
	3. Environmental
	4. Failure of preventive/protective equipment
Correct hazards	2, 4
What are the possible consequences of this risk?	a. Risk of explosion
	b. Insufficient ventilation of highly toxic fumes, causing inhalation problems
	c. Local ventilation can ignite the chemical, causing fire
	d. Glass will break causing a spill and corrosive burn injury
Correct consequences	b
Hint	I smell a pungent smell of garlic or onion
Safety information	Highly toxic chemicals should be in a well ventilated area. Local ventilation is not sufficient
Correct procedure	Move the flask to a fume hood and turn on ventilation

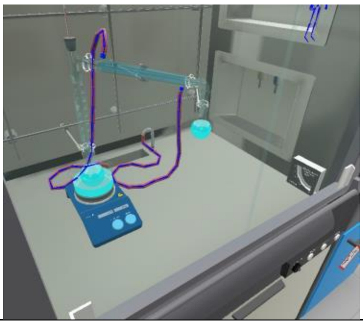
Risk 11: Laboratory coat in the office room

Location	Office room
Items	Lab coat on a chair
Which hazard types are present in this photo?	1. Physical
	2. Health
	3. Environmental
	4. Failure of preventive/protective equipment
Correct hazards	2
What are the possible consequences of this risk?	a. The lab coat will catch on fire
	b. The office will smell bad because of the lab coat
	c. The lab coat is contaminated by chemicals, risking the health of people
	d. The lab coat will fall on the ground causing tripping hazard
Correct consequences	c
Hint	There is something that does not belong in the office
Safety information	Never bring any laboratory equipment in the offices. It could contaminate non-laboratory environments with hazardous chemicals
Correct procedure	Move lab coat to the PPE room

Risk 12: Fume hood window too much open

Location	Fume hood
Items	Synthesis: nylon-11 monomer. Beaker and bottle of ammonia, bottle of 11-Bromodecanoic acid
Which hazard types are present in this photo?	1. Physical
	2. Health
	3. Environmental
	4. Failure of preventive/protective equipment
Correct hazards	2, 4
What are the possible consequences of this risk?	a. The hot plate can ignite chemicals on fire
	b. Insufficient ventilation causes inhalation problems
	c. Chemical splashes can reach the body easily
	d. Insufficient ventilation can create an explosive atmosphere
Correct consequences	b, c
Hint	A sharp odour hurts my nose!
Safety information	Fume hood window should be lower in order to protect yourself from splashes and to increase ventilation power.
Correct procedure	Pull fume hood window down

Risk 13: Fume hood ventilation not turned on



Location	Fume hood
Items	Distillation setup: synthesis of ethyl mercaptan, ventilation is off. Distillate: ethyl mercaptan. Residue: sodium ethyl sulphate, potassium hydrosulphide
Which hazard types are present in this photo?	1. Physical 2. Health 3. Environmental 4. Failure of preventive/protective equipment
Correct hazards	1, 4
What are the possible consequences of this risk?	a. Insufficient ventilation can create explosive atmosphere b. Chemical splashes can reach the body easily c. No condensation of vapours can create explosive atmosphere d. Toxic fumes can cause inhalation problems
Correct consequences	a
Hint	All ventilation systems need to be checked frequently
Safety information	Flammable gases and vapours can ignite easily or build up to explosive atmospheres when ventilation is insufficient.
Correct procedure	Turn the ventilation of the fume hood on.

Behavioural risks in the laboratory

Behavioural risks are risks that appear due to the behaviour of a character in the game. A non- player character (NPC) can show this at-risk behaviour via their appearance or their actions played as an animation

Behavioural risk 1: Absence of correct PPE

Behaviour/appearance	NPC appearance
Items	Normal clothing, no lab glasses, no gloves, no lab coat
Which hazard types are present in this photo?	1. Not wearing eye protection 2. Incorrect casual clothing 3. Not wearing lab coat 4. Incorrect hairstyle
Correct hazards	1, 3
What are the possible consequences of this risk?	a. Violation of the laboratory rules b. Injury to the skin in case of incident c. Hair can be affected in an incident d. Eye irritation or blindness if incident happens
Correct consequences	a, b, d
Hint	"Wearing PPE is annoying, I only need to be in the lab for a short time"

Safety information	To prevent risk of injury when there is an incident, it is obligated to wear lab glasses and lab coat in the lab
Correct procedure	Give lab glasses, lab gloves and lab coat to the NPC

Behavioural risk 2: Improper length of trousers

Behaviour/appearance	NPC appearance
Items	Shorts or skirt with short length, lab coat, lab glasses, nitrile gloves
Which hazard types are present in this photo?	1. Not wearing eye protection
	2. Incorrect casual clothing
	3. Not wearing lab coat
	4. Incorrect hairstyle
Correct hazards	2
What are the possible consequences of this risk?	a. Violation of the laboratory rules
	b. Injury to the skin in case of incident
	c. Hair can be affected in an incident
	d. Eye irritation or blindness if incident happens
Correct consequences	a, b

Hint	The lab can get really hot in the summer. Better to wear more comfortable clothing
Safety information	To protect your skin, always wear clothing with long sleeves and long trouser length to cover exposed parts of the body
Correct procedure	Give long trousers to the NPC. Remove lab coat first and give it back

Behavioural risk 3: Improper length of shirt

Behaviour/appearance	NPC appearance
Items	T-shirt with short sleeves, no lab coat, lab glasses, nitrile gloves
Which hazard types are present in this photo?	1. Not wearing eye protection
	2. Incorrect casual clothing
	3. Not wearing lab coat
	4. Incorrect hairstyle
Correct hazards	2, 3
What are the possible consequences of this risk?	a. Violation of the laboratory rules
	b. Injury to the skin in case of incident
	c. Hair can be affected in an incident
	d. Eye irritation or blindness if incident happens
Correct consequences	b

Hint	The lab can get really hot in the summer. Better to wear more comfortable clothing
Safety information	To protect your skin, always wear clothing with long sleeves and long trouser length to cover exposed parts of the body
Correct procedure	Give shirt or sweater with long sleeves to the NPC

Behavioural risk 4: Incorrect hairstyle

Behaviour/appearance	NPC appearance
Items	Long hairstyle, Lab coat, nitrile gloves, lab glasses.
Which hazard types are present in this photo?	1. Not wearing eye protection
	2. Incorrect casual clothing
	3. Not wearing lab coat
	4. Incorrect hairstyle
Correct hazards	4
What are the possible consequences of this risk?	a. Violation of the laboratory rules
	b. Injury to the skin in case of incident
	c. Hair can be affected in an incident
	d. Eye irritation or blindness if incident happens
Correct consequences	a, c
Hint	In the lab, personal protection is more important than personal look.
Safety information	Long hairstyle can come in contact with chemicals more easily. So long hair must be tied back in the lab.
Correct procedure	Give hair band to the NPC

Behavioural risk 5: Mouth pipetting

Behaviour/appearance	NPC transfers liquid chemical with his/her mouth
Items	Volumetric pipet
Which hazard types are present in this photo?	1. Incorrect personal protection equipment (PPE)
	2. Too much pipette volume
	3. Dangerous pipetting technique
	4. Not environmental friendly
Correct hazards	3
What are the possible consequences of this risk?	a. Possible ingestion of chemicals
	b. Pipetting technique will suffocate this person
	c. Chemicals are exposed to air
	d. Pipetting incorrect volume
Correct consequences	a
Hint	In the good old days, mouth pipetting was the only way to transfer liquids
Safety information	Mouth pipetting increases chance of ingesting chemicals. Use a pipette filler or air displacement pipettes instead.
Correct procedure	Give pipette filler or air displacement pipette

Appendix 7.B: MCQs for MAR lab game

	QUESTIONS:
1	Should it be the summative or the ranges per metric?
2	Do you think that somehow, we can incorporate the #data points and time during titration? If yes, how would you recommend it?
3	Can we check the ranges?

Appendix 7.C: MCQs for CosmiClean

Pre-game tests

1. A magnetic separator is best suited for removing which of the following *
    a. Ferrous metals
    b. Non-ferrous metal
    c. Glass
    d. Water

2. A processor that attracts non-ferrous metals will remove which of the following? *
    a. Copper
    b. Iron
    c. Wood
    d. Glass

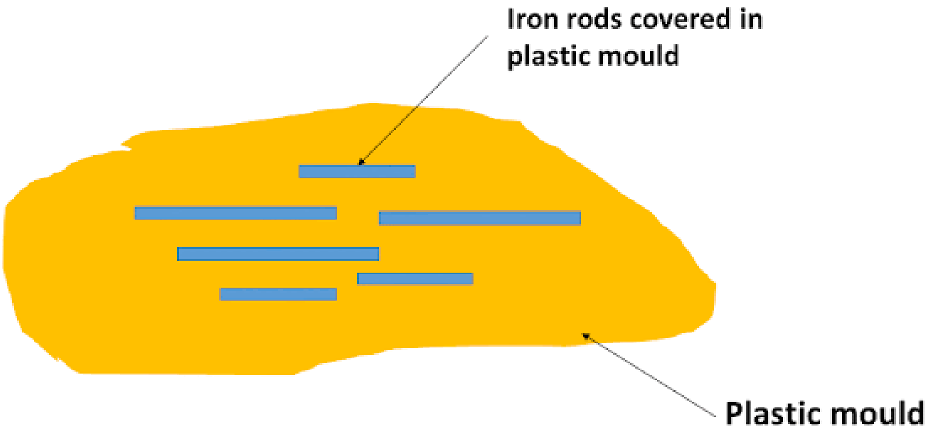
3. Which of the following mixtures can be separated with a boiler? Select all applicable.
    a. Salt solution
    b. Water and oil
    c. Glass and plastics
    d. Plastic and Iron

4. Which of the following would be easily separated by adding a solvent? Select all applicable.
    a. Sand and salt
    b. Plastic and glass
    c. Sugar and flour
    d. Iron and concrete

5. Which of these waste materials will be best separated with a sieve?
    a. Plastic bottles (5cm) and metal rods (5mm)
    b. Wood chips(2mm) and pieces of glass (2mm)
    c. Iron rods (3mm) and wood (3mm)
    d. Water and petrol (2 litres)

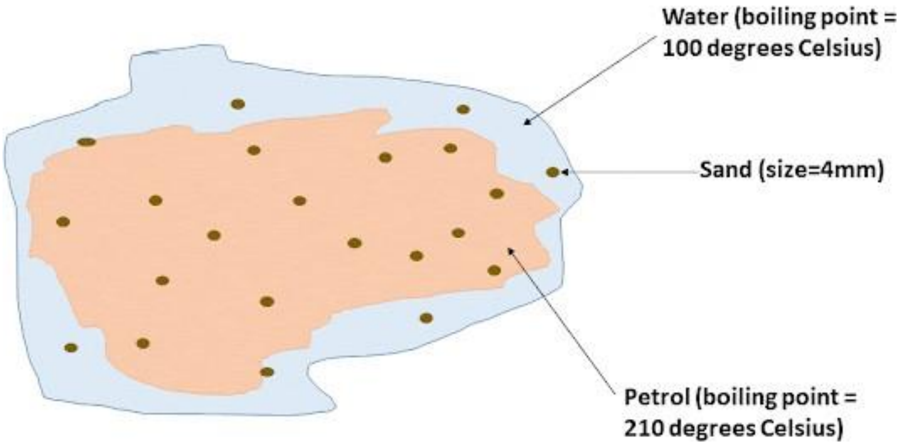
6. A vanload of waste from a burnt warehouse was delivered for separation as shown below.

    How would you separate these materials without using heat?



a. Shredding => Magnet
b. Magnet => Shredding
c. Magnet => Sieving
d. Melting => Sieving

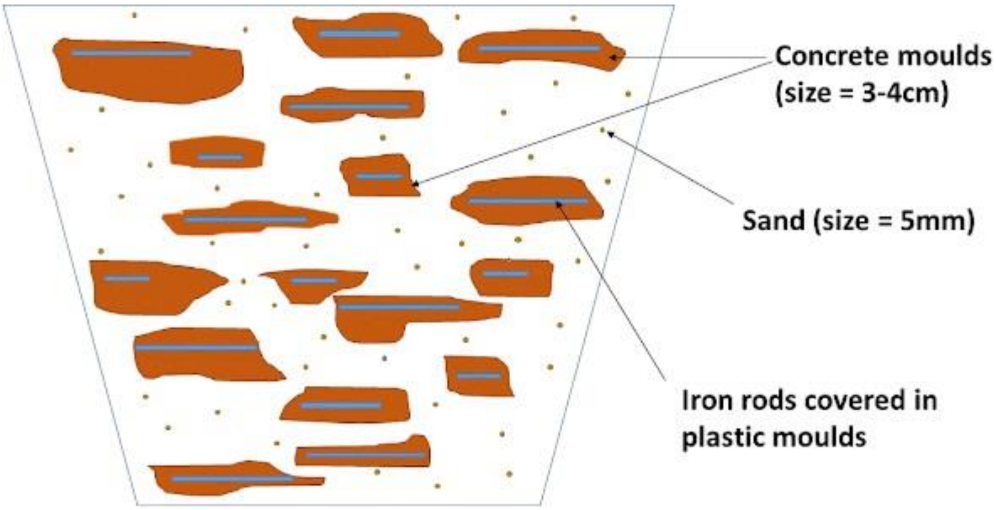
7. What are the most suitable processors for separating the following? *



a. Boiler => boiler
b. Shredder => boiler
c. Sieve => boiler
d. Boiler => shredder

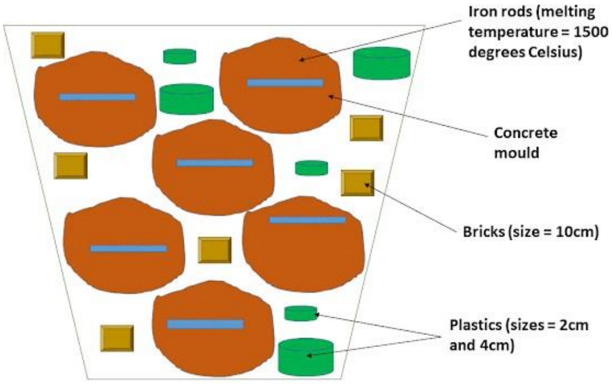
8. An alloy of copper and gold arrived for separation. The melting temperature of both gold and copper is 1050 degrees Celsius. How would you consider separating the copper from gold?
   a. By dissolving
   b. By melting
   c. By shredding
   d. By boiling

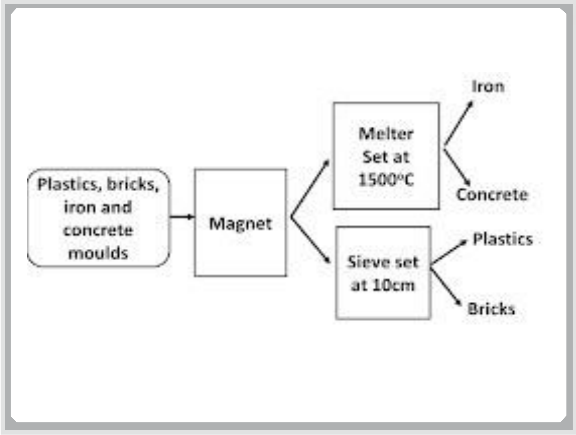
9. Which of the following sequences will efficiently separate these waste materials into their individual particles?

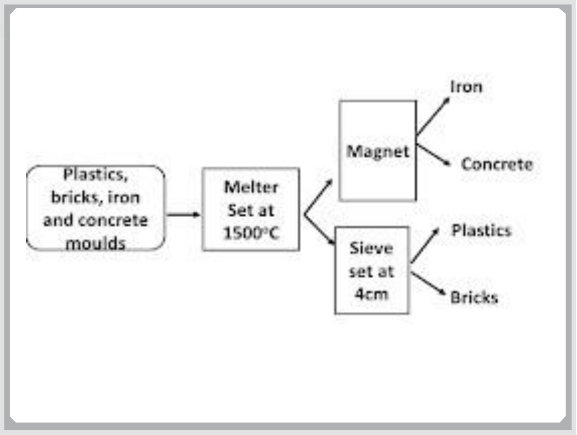


a. Sieve => Melter
b. Sieve => Sieve
c. Melter => Sieve
d. Melter => Melter

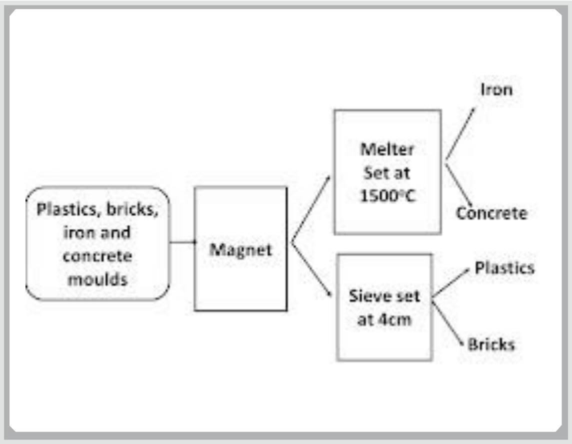
10. How would you efficiently separate the following?

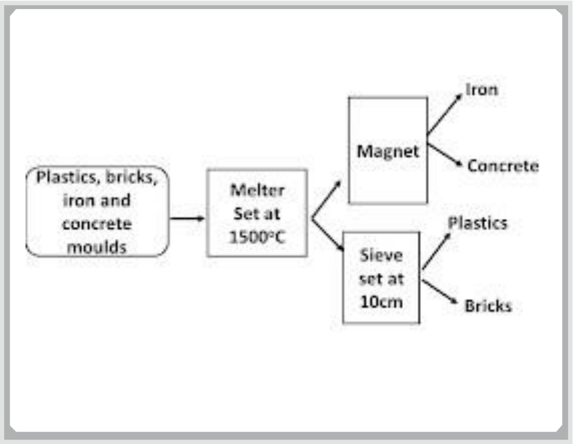






Option 1 Option 2



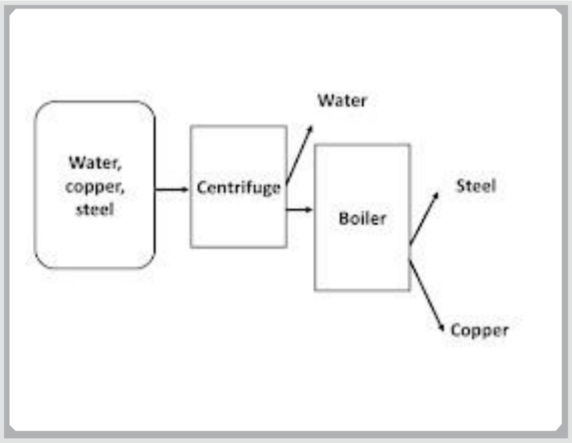


Option 3 Option 4

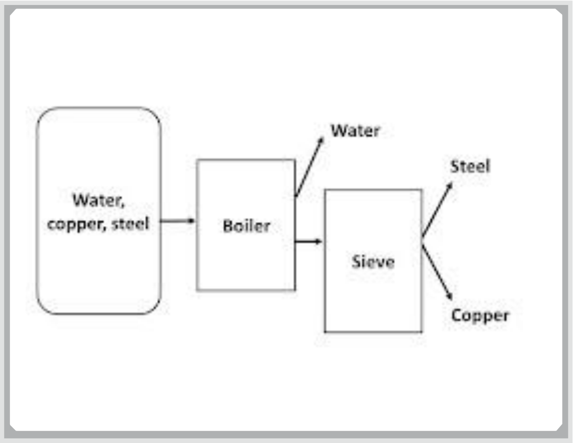
11. Given waste materials containing copper, plastics, and iron of sizes 4cm each, what two processors will you use to separate them?
    a. Magnet => Eddy current separator
    b. Sieve => Boiler
    c. Melter => Sieve
    d. Eddy current separator => Sieve

12. You have a sieve, a boiler, and a centrifuge. How would you separate wastes containing water, copper, and steel?

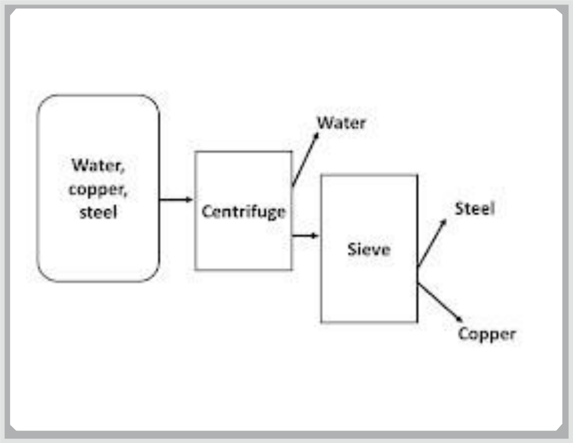
MATERIALS	PROPERTIES
Water	Density is 1g/cm3 Boiling temperature is 100 degrees Celsius
Copper	Density is 8.9g/cm3 Size is 2mm
Steel	Density is 7.8 g/cm3 Size is 2mm



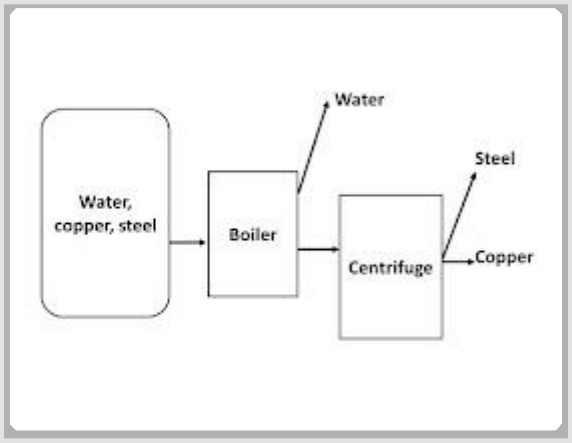
Option 1



Option 2

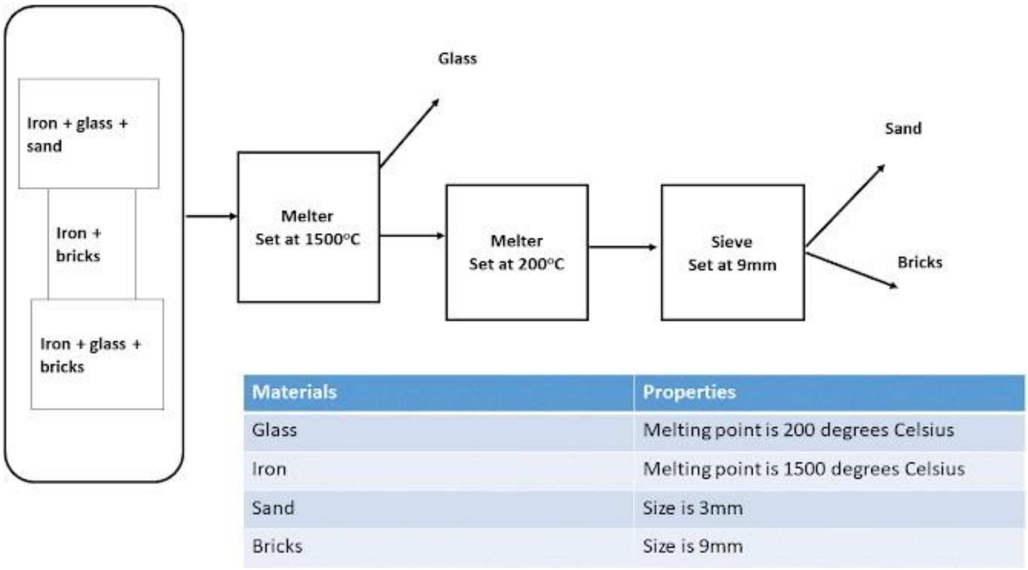


Option 3



Option 4

13. An intern at your factory made some mistakes when separating glass, iron, sand, and bricks. Looking at the plan below, what went wrong and how would you rectify the mistakes?



Materials	Properties
Glass	Melting point is 200 degrees Celsius
Iron	Melting point is 1500 degrees Celsius
Sand	Size is 3mm
Bricks	Size is 9mm

Post-game tests

1. Which of the following processes materials based on temperature?
    a. The melter
    b. The dissolver
    c. Centrifuge
    d. Magnetic separator

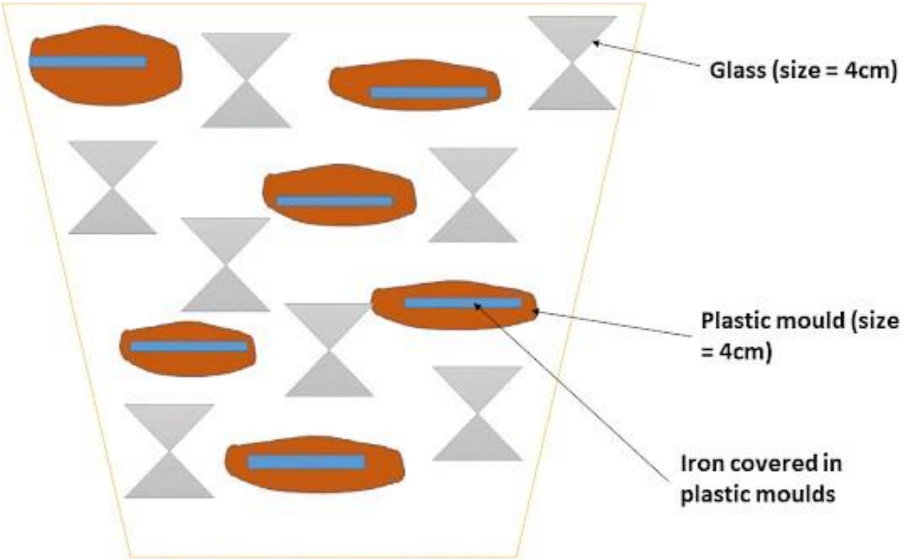
2. A shredder processes materials by
    a. Size reduction
    b. Melting
    c. Dissolving
    d. Boiling

3. Which of the following mixtures can be separated with a boiler? Select all applicable. *
    a. Water and Petrol
    b. Sand and Plastic
    c. Ethanol and Water
    d. Iron and Copper

4. A dissolver relies on a material's solubility while a centrifuge relies on a material's *
    a. Density
    b. Boiling point
    c. Magnetic properties
    d. Melting point

5. A magnetic separator will be most suited for separating which of the following waste materials? Select all applicable.
    a. Iron and wood
    b. Sand and concrete
    c. Steel and glass
    d. Copper and plastic

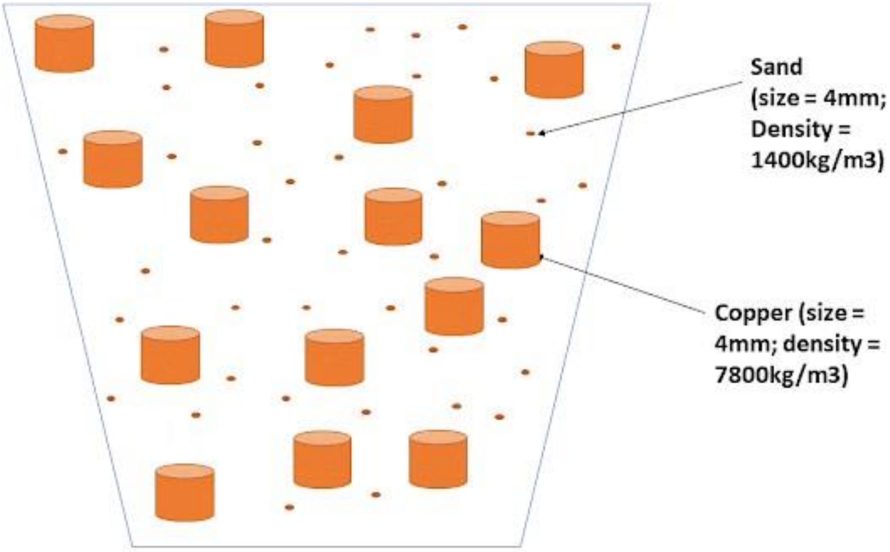
6. How would you separate the glass, plastics, and iron? *



a. Eddy current separator, Shredder
b. Magnetic separator, Shredder
c. Sieve, Magnetic separator
d. Shredder, Sieve

7. Glass and iron wastes are most suited for a magnet while salt and water mixtures will be best suited for
   a. Boiler
   b. Shredder
   c. Melter
   d. Eddy current separator

8. Which of these would you use to separate the waste below given the information provided?



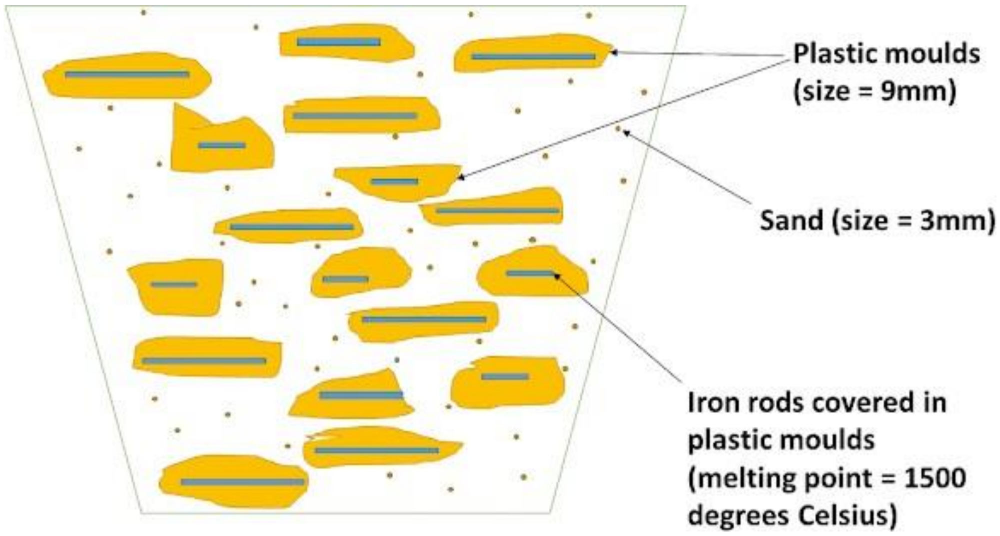
a. Centrifuge
b. Magnetic separator

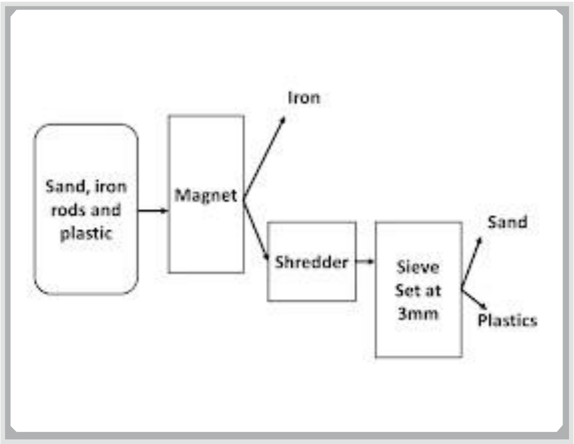
c. Sieve
d. Melter

9. Given wastes containing sands of size 4mm, woods of size 9mm, and wine corks of size 12mm. How would you efficiently separate these?

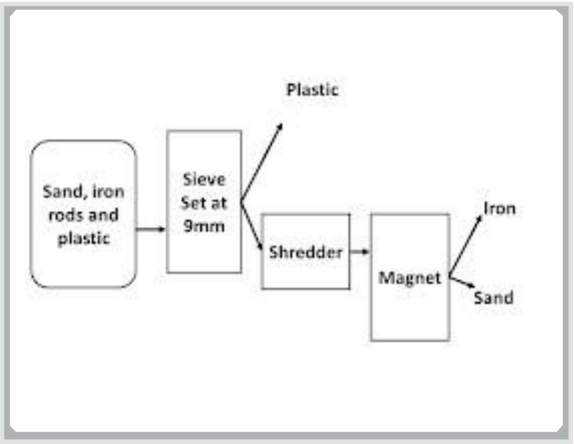
   a. Sieve mesh set at 9mm, and sieve mesh set at 4mm
   b. Sieve mesh set at 12mm, and sieve mesh set at 9mm
   c. Sieve mesh set at 9mm, and sieve mesh set at 12mm
   d. Sieve mesh set at 4mm, and sieve mesh set at 12mm

10. Waste from a burnt down warehouse made up of sand, iron rods and molten plastics was delivered for separation. How would you separate these?

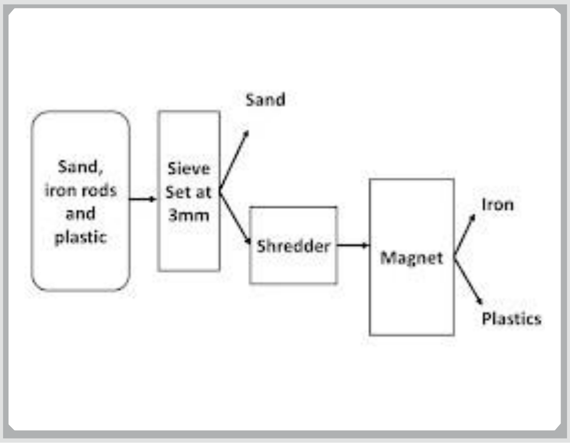




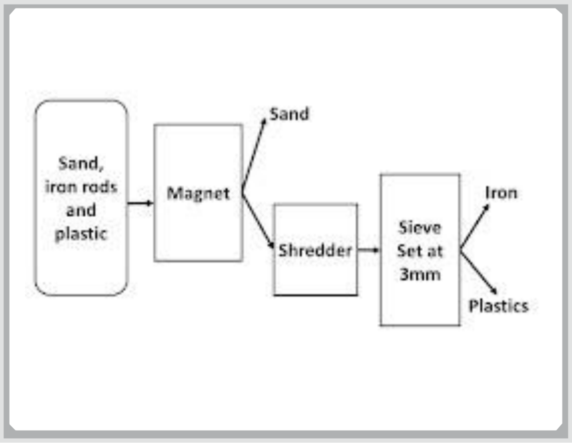
Option



Option 2



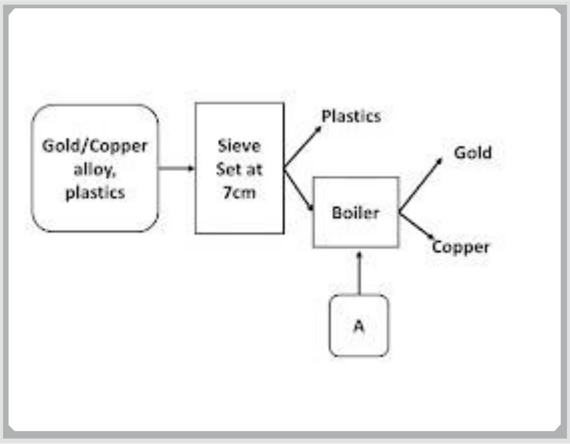
Option 3



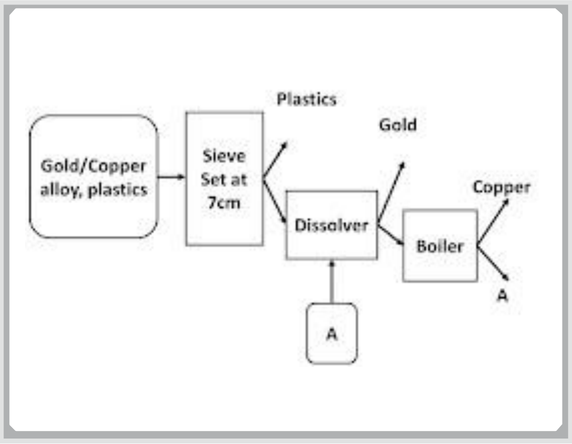
Option 4

11. Waste consisting of water, iron and plastics in the ratio of 1:2:1 will be best separated using
    a. Boiler => Magnetic separator
    b. Magnetic separator => Sieve
    c. Magnetic separator => Shredder
    d. Shredder => Boiler

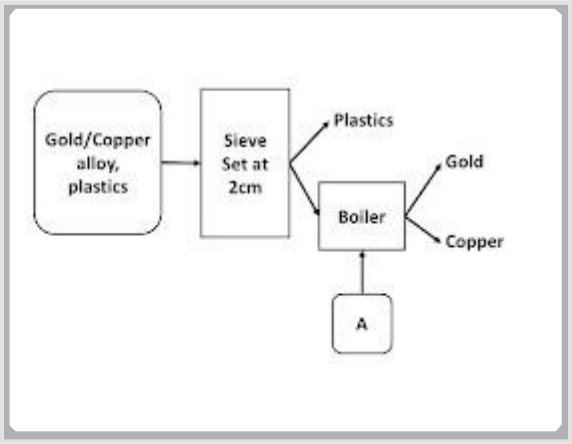
12. Gold and copper alloys of sizes 7cm are mixed with plastics of sizes 2cm. How would you separate these knowing that copper is soluble in solvent "A"?



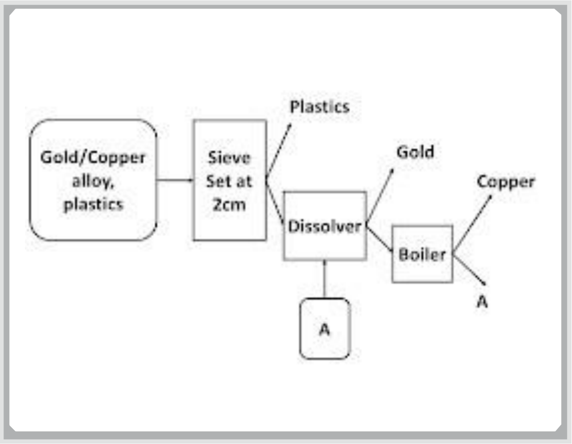
Option 1



Option 2

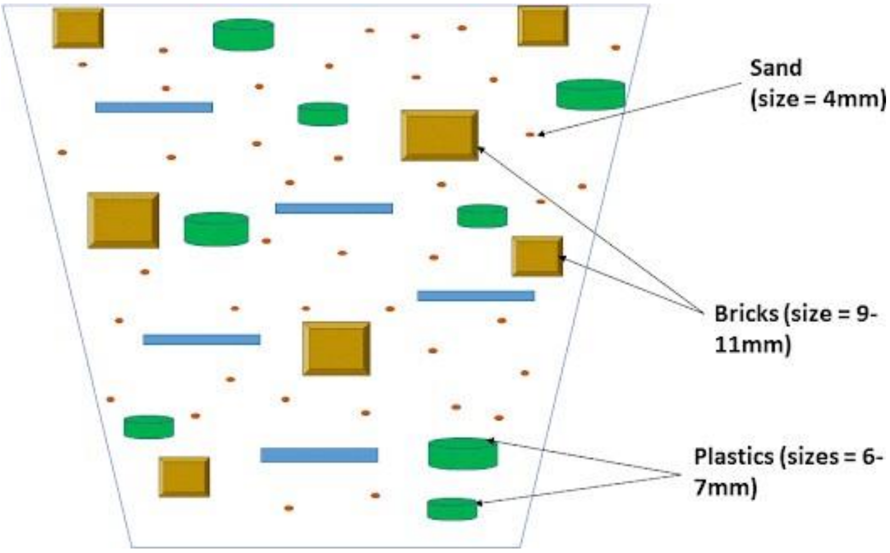


Option 3

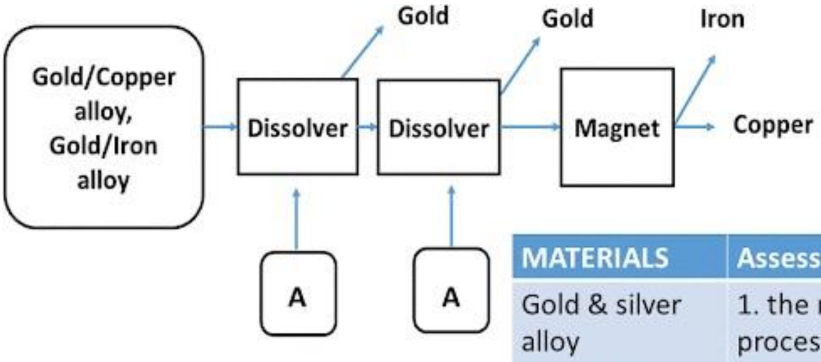


Option 4

13. Given a waste material consisting of sand, plastics, and bricks, what sequence and configuration would you use to separate these given the information provided? Discuss the reasoning behind your solution.



14. As a senior process engineer, you asked a trainee to describe how he would separate 200kg of a mix of gold & silver alloy and gold & iron alloy. An hour later, he presented you with the illustration below. Assess this plan based on the criteria provided and complete the task.



MATERIALS	Assessment Criteria
Gold & silver alloy Gold & iron alloy	1. the numbers and types of processors used 2. the sequence of processor 3. materials extracted

Your Task: To make corrections to the above plan, describe what the correct sequence of the processors should be and outline where each material will be extracted.

NB: Iron is soluble in "A".