Minimising Extraneous Cognitive Load in Immersive Virtual Environments: Evaluating an Immersive Virtual Reality Educational Platform Against the Principles of Cognitive Load Theory

Teaching Dissertation submitted in partial fulfilment of the requirements for the degree of

Masters of Teaching

(EDUC 7557)

Jock A.M. Rogers

BSc

Research Mentor/Supervisor

Dr Brendan Bentley

The University of Adelaide

Due: 7th June, 2019

Submitted: 12th June, 2019

Word Count (Introduction-Recommendations): 26,640

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

Signed: _______________________________

Name:

Date:

Abstract

Many consider STEM skills to be increasingly important for the future workplace. However, Australian enrolments in senior secondary STEM subjects declined through the 90s and early 2000s, especially in science subjects. The number of enrolments plateaued and has changed little since. One reasons behind this is that students tend to avoid and dropout of STEM subjects. This is partly because STEM education across Australia is not fostering enough interest in students, and a lack of engagement and enjoyment in the classroom. STEM Classrooms need to be made more interesting, engaging and enjoyable, but in a way that does not harm learning. Immersive Virtual Reality (IVR) is uniquely positioned for this as it can create novel, authentic, immersive, interactive and emotional experiences. These experiences immerse users within the virtual environment (VE), establishing a sense of presence or 'being there'. Presence and authenticity create unmediated and engaging experiences, of the type recommended by most modern learning theories. However, IVR comes with many limitations, of which cost is the most notable. Furthermore, there are significant gaps in the literature describing and demonstrating the relation between the countless factors that define IVR, and learning outcomes. So far studies have demonstrated that IVR is often not superior, and sometimes inferior, to traditional methods with respect to cognitive learning outcomes. A problem which is exacerbated by the fact that few of the many IVR devices and software developed are grounded in solid pedagogy. Early research has indicated why this might be the case, the answer might come from Cognitive Load Theory (CLT). CLT is a learning theory especially suited for describing the cognitive loads associated with learning tasks and the methods to manage and reduce it. IVR is often informationally dense, requiring that students navigate a full 360˚, 3D virtual environment whilst being dazzled by many sources of visual and sometimes audial information. CLT suggests that cognitive load comes in two forms. One of these, extraneous load, encompasses content that is irrelevant to learning that either distracts students, or forces them to process it alongside task-relevant information. IVR, it is suggested, tends to create large extraneous loads, which is possibly the source of its mediocre performance in producing cognitive learning outcomes. The goal of this study was to use the principles of CLT to evaluate an educational IVR platform with regards to minimising extraneous load. This platform could support large numbers of users simultaneously within shared or separate VEs whilst an instructor, using a separate non-IVR device, selected, controlled and manipulate the content. After a review of the literature, two questions would define the evaluation: Q1) How does the educational IVR platform compare to the recommendations of CLT regarding the reduction of extraneous load in the presentation of content? Q2) Where & how could the educational IVR platform be changed to better meet the recommendations? The evaluation was conducted by directly using the platform, during which the platform’s content was qualitatively observed and its characteristics explored. These observations were conducted using a coding framework consisting of criteria that was synthesised from the principles of CLT and the characteristics, capabilities and limitations of the platform initially identified. The observations were analysed and discussed, in these discussions content design methods based off the criteria were suggested. More importantly, the interaction between the defining characteristics of the platform and the criteria was determined, from which the primary evaluation and recommendations were made. The evaluation concluded that the platform was well suited for minimising extraneous load for several reasons (Q1). Some simple recommendations were made, primarily the addition of more tools for the instructor to use to manipulate running content (Q2). By attempting to evaluate an ICT technology, the research aimed to guide or assist future evaluations. However, due to limitations in the literature and research method, assumptions had to be made. The limitations that necessitated these assumptions provided a basis for suggested research directions.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to state my earnest gratitude for my supervisor and program director Dr Brendan Bentley, for guiding me through this project and giving me the confidence to see it through. He was very patience with me, providing extra time to complete the dissertation since my opportunity to collect data came late. I would also like to thank Dr Lynda Macleod for not only coordinating the course but also her involvement with the legal organisation associated with my thesis.

I must thank the Curiious staff for affording me the opportunity to work with their exciting platform. They made multiple long trips from NSW to SA to demonstrate the platform to the department and later gave me personal time analyse it. I believe they are making a rare and earnest effort to design a system with the wisdom of educational research and theory in mind. I wish them all the best for the future and I hope to see their device appear in classrooms soon.

I would like to thank my peer Di Hu for our conversations regarding the course and our topics, and for being prepared to check my project even though I had missed the opportunity to organise it. She is experienced with research and completed an excellent dissertation, so on more than a few occasions she shared some helpful advice with me. I would like to acknowledge all of my fellow Masters students, those involved with organising the course and the Faculty of Arts in general. I have felt welcomed and supported throughout this course.

Finally, I would like to acknowledge my family: my older brother Hunter and twin brother Fergus, my parents Judith and Timothy, and Jack, my old dog, who unfortunately passed away near the end of this dissertation. Fergus has been through research before and is currently completing a PHD, he is far better versed in academia than I am and supported me during the project. My parents put in the herculean effort of raising my brothers and I, they supported me during this project and I have the immense comfort and privilege of knowing that they are there for me when I need them.

1 Introduction

1.1 The Problem

"…the availability of people with business and entrepreneurial skills and skills in science, technology, engineering and maths (STEM) seem to be universally accepted as necessary skills for the future" (Australian Government. Productivity Commission, 2016). As technology continues to become more integrated with our lives, a workforce that is skilled in STEM is considered to be extremely important to the future economic prosperity of Australia (Panizzon, Corrigan, Forgasz, & Hopkins, 2015). This is due to the technologies reliance on STEM based skills, itself changing the way STEM is practiced. In a study by the Office of the Chief Scientist, most employers considered STEM qualifications as valuable, considered STEM-skilled employees the most innovative, and expected their needs for STEM-skilled professionals to increase in the near future (Prinsley & Baranyai, 2013).

Despite the importance of STEM skills, the rates of enrolment in senior secondary STEM subjects declined through 1990 to 2000, and plateaued since (Panizzon et al., 2015). Furthermore, reports state that students underperform in these subjects, that jobs requiring STEM skills are growing faster than other areas of employment and many employers have difficulties recruiting workers with STEM qualifications (Newhouse, 2017; Prinsley & Baranyai, 2013). Students tend to readily drop out of STEM studies or avoid them altogether, especially at the senior secondary level. This problem of student participation is partly responsible for the decline in enrolments (Newhouse, 2017). Understanding the sources of student dropout will be a step forward in filling the growing vacancies in STEM careers and maintaining Australia's competitiveness in the global economy. Regardless of the reasons, part of the solution undoubtably involves the training of teachers in effective teaching practice (Newhouse, 2017), that its, teaching through means and methods that encourage students to enrol in STEM subjects and see them through.

There are numerous possible reasons for the decline in STEM enrolments and rise in dropouts (Newhouse, 2017). Studies report that students find STEM subjects to be challenging, boring and uninteresting. Furthermore, the STEM curriculum might be too focused on large bodies of factual and procedural knowledge, and too lacking in cross-disciplinary activities, creativity and collaboration. The general message from studies that examine the issue is that no single factor that influences STEM enrolment is universal, they are context dependent, intertwined and often difficult to untangle (Regan & DeWitt, 2015). Amongst these factors are students' personal interest in STEM subjects (as a psychological construct in choice), and their experience learning STEM in schools (teaching methods and quality). Interest is a complex psychological construct, but is thought to be composed of a cognitive and affective component and sustained by interaction. Interest in science is a highly influential factor in choosing to study science, identified as the dominant enrolment influence in studies examining it. Interest has strong predictive relations with personal value, enjoyment, and intention of learning science. Teaching quality is a “major determinant of student engagement with and success in a school subject” (Regan & DeWitt, 2015, p. 72). Teaching quality (as well as curriculum) can affect the student experience in several ways. Studies suggest that students experience a poor transition from primary to secondary schooling, low levels of experimental work in class, lack of purpose and appeal in their classes, and struggle to imagine themselves in science, to name just a few issues. Undoubtably, a significant responsibility rests on STEM teachers to adjust their methods to become more effective teachers. Amongst other things, effective teaching practice includes the employment of methods and technologies that make science more interesting, engaging, enjoyable and less difficult. STEM education itself is about more than the subjects that compose it, it is an “integrative approach to curriculum and instruction” (A. Roberts, 2012). For secondary education, an embedded curriculum is probably the best approach to improving STEM outcomes: “situated approaches that emphasize the learning of domain knowledge through expert-like activities and authentic problem solving in rich social, cultural, and functional contexts” (Chen, 2001, p. 194). Engaging in such approaches should be a step forward in improving the STEM education experience and fostering student interest in its subjects.

1.2 Setting the Scene

Part of the solution to STEM dropout and retention comes with the effective implementation of ICT. The capability of ICT to support better learning environments and activities has long been noted (Newhouse, 2017). ICT can support active engagement, focus attention to particular aspects of learning, relieve students from tedious work and support more interesting activities. However, to this end it is best to apply ICT in the support of learning, to use ICT with the intention of largely replacing traditional teaching methods is ill advised. It is generally accepted that there is no direct link between using ICT and the production of better learning outcomes in the classroom (Higgins, 2003; Lei & Zhao, 2007; Voogt, Knezek, Cox, Knezek, & ten Brummelhuis, 2013). Rather, it is how ICT broadens the capabilities of a teacher, such that they can create better learning environments, that makes the difference. Hence, insofar as ICT can help, the technologies available and how teachers implement them in the classroom will dictate whether an activity, class or program can be improved and ultimately lead to higher enrolment, lower attrition and better learning outcomes for STEM.

The effects of ICT in the classroom have been extensively researched. These effects include, with the correct implementation, better outcomes in student engagement (Passey, Rogers, Machell, McHugh, & Allaway, 2004), higher order thinking (Lim & Tay, 2003), and motivation (Zhang, 2008), to name a few. A literature review by Newhouse (2015) suggests that ICT can improve attributes of learning environments in the following categories: “Investigating real problems and data; Building knowledge; Promoting active learning; Supporting authentic assessment; Engaging students by motivation and challenge; Providing tools to increase student productivity; Providing scaffolding to support higher level thinking; Increasing learner independence; Increasing collaboration and cooperation; Tailoring learning to the learner; and Overcoming physical disabilities." Overall, ICT provides the ideal tools for constructing an embedded STEM curriculum that will allow students to get used to the types of technologies that may be central to their future workplaces.

Virtual reality (VR), facilitated by headsets and other fully enclosing devices, is a type of ICT that is uniquely positioned to create engaging and motivating environments. It can transport concepts from the abstract descriptions of traditional teaching to concrete and interactable forms that unlock the potential of different learning styles, and foster affective outcomes by triggering deeper emotional experiences than other mediums. VR is often implemented with the belief that its novelty and capability to create an authentic experience will draw interest from students, and then accommodate interactive learning within context relevant environments (Kavanagh, Luxton-Reilly, Wuensche, & Plimmer, 2017; Markowitz, Laha, Perone, Pea, & Bailenson, 2018). Therefore, the unique affordances of VR could play a key role in reimagining STEM education by reinvigorating interest, curtailing boredom, promoting deeper understanding and facilitating more natural learning environments. Following the trend of ICT in general, VR is not a silver bullet, the implementation of it alone will not lead to better educational outcomes for STEM or otherwise. It is safe to assume that VR can only help if it is implemented correctly. Studies note that many teachers lack pedagogical competency with ICT in general (often summarised as ‘technological pedagogical content knowledge’ or TPACK) (Voogt et al., 2013) let alone with VR. Enlarging this problem is the fact that VR is neither cheap nor easily employed (Kavanagh et al., 2017), requiring adequate training for instructors & users and the development of supporting school policies and procedures. If it is to play a part in increasing interest and learning in STEM education it is critical, therefore, to design VR software and hardware that is congruent with educational theory and classroom reality.

1.3 Rational

This study was interested in improving student cognitive learning outcomes when using the platform (Anderson, Krathwohl, Airasian, & Bloom, 2001: Revised Bloom's Taxonomy), and in a way that does not significantly change its affective potential given the unique affective affordances of VR. It seems obvious that VR can make learning more interesting, but whether learning outcomes themselves are improved is a different question. As previously discussed, the objective of improving STEM outcomes includes increasing enrolment and retainment, but also the improvement (and not at the detriment) of learning outcomes.

Learning outcomes are dependent on different dimensions of cognitive theories and pedagogy, to focus this study only the content presentation of the platform (the design of learning tasks) would be evaluated. *Cognitive Load Theory* (CLT) provided the ideal theoretical basis for this objective given the focus on content presentation, as it was “explicitly developed as a theory of instructional design based on our knowledge of human cognitive architecture” (Sweller, Ayres, & Kalyuga, 2011, p. V). Consequently, the basis of evaluation was developed entirely from an integration of the principles of CLT and the specific affordances of VR, namely, the characteristics of the platform. CLT and related theories are described in detail within the literature review (2.3-2.4), and its application further justified in the methodology (3.2). Ideas from most educational theories (e.g. constructivism, discovery learning, situated cognition etc.) that describe the types of activities and instruction that are worthwhile (whether to show the class a video or have them play a game, whether to lecture students or create a discussion, and so on) were not considered by this study unless crossover was necessary.

The aim of this project is best summarised as an attempt to apply the recommendations of an educational theory to evaluate a VR platform. More specifically, the VR platform was evaluated against criteria synthesised from the recommendations of CLT and characteristics of the platform. Before this could be done, an overview of the literature and current problems was necessary (2), in particular, research that measured or discussed cognitive load in VR (2.5.5). The methodology describes the design and details of the study (3), followed by the results & discussion in which the data was analysed & discussed and the platform’s evaluation is made (4). After the evaluation, conclusions, and recommendations for future research directions and future investigators were made (5**Error! Reference source not found.**), primarily based on the literature review and the e xperience of the study. Some immediate limitations of the adopted approach were that: it did not include primary research with school students (the target audience of VR); it worked with theory garnered from a literature that was still in its infancy. Nonetheless, a broad analysis was to be made, using extensive descriptions which would leave ample room to discuss alternative ideas and exceptions.

2 Literature Review

2.1 Introduction

This literature review explores the virtual reality literature and the theories of working memory and cognitive load. The properties of IVR that are thought to influence learning are described, and the theories behind why they do are explained. A general overview of the empirical research is conducted, from which the relationship between virtual reality and cognitive load theory is determined. Research into the educational effects of VR is new, which can be seen in the range of research goals and lack of consistency in nomenclature. As such this review casted a relatively wide net, from which an effort was made to identify the common threads.

2.2 Immersive Virtual Reality

VR is defined as “(minimally) a digital representation of a three dimensional object and/or environment” (Kavanagh et al., 2017, p. 86). The working definition of VR varies within the literature as the minimal definition includes anything from 2D screens to fully enclosed environments (Johnson-Glenberg, 2018; Kavanagh et al., 2017). As a result, a distinction has been established in the of *Immersive VR* and *Non-Immersive VR*. Immersive VR (IVR) “systematically maintains an illusion of presence, such that learners feel their bodies are inside the virtual environment” (Johnson-Glenberg, 2018, p. 2). These systems create an immersive 3D experience where an engaged user cannot see the real world for 360˚, being at the very least visually contained within a virtual environment (VE) (Jensen & Konradsen, 2018). Using the minimal definition of VR, non-immersive VR encompasses the complement of IVR, typically represented by regular 2D displays. There has been growing interest in IVR as of late, with some researchers predicting that it will accompany a paradigm shift in multiple fields, including education. On one hand, large companies like Google, Apple, Facebook and many others are driving the development of IVR hardware and software (Makransky, Terkildsen, & Mayer, 2019). On the other hand, the novel and unique capabilities of IVR are central to the beliefs in its educational capabilities, especially its capability of creating interactive and realistic environments, which is driving the education based research effort (Kavanagh et al., 2017; Markowitz et al., 2018). To begin with, an overview of IVR including its history, affordances & capabilities, applications and issues is outlined in sections 2.2.1-2.2.4.

2.2.1 History

IVR was first implemented for an educational purpose in the 1960s with a flight simulator developed for the United States air force. IVR remained mostly in the public sector until the 90s which saw the release of IVR arcade games, however these were not popular and were discontinued soon after release. Following this both SEGA and Nintendo developed head mounted display based IVR games.

SEGA's never released and Nintendo's was a commercial failure (Freina & Ott, 2015; Kavanagh et al., 2017). The 90s also saw the application of VR in STEM education based off positive beliefs in its motivating and authentic capabilities. Eventually many VEs were developed for non-IVR environments, consistent with the classic example of 'River City' (Dede, 2009) where students would develop scientific research questions regarding the city's issues (Markowitz et al., 2018). Most applications in education were limited to non-IVR due to the impractical costs of the IVR technology and its technological limitations (low resolution, poor image latency, poor head-tracking accuracy etc. Kavanagh et al. (2017)). IVR saw a resurgence in the 2010s with the release of the Oculus Rift and related devices (Jensen & Konradsen, 2018; Johnson-Glenberg, 2018; Southgate et al., 2019), these overcame many of the issues afflicting the older hardware such as accuracy, usability, price and consumer availability (Johnston, Olivas, Steele, Smith, & Bailey, 2018; Kavanagh et al., 2017). Recently, VR educational research is becoming increasingly common (Makransky et al., 2019), Vergara et al. (2017) found that the number of VR studies per year has been steadily increasing since the 90s. In a review of the medical training VR literature by Kyaw et al. (2019), no papers were found before 2005 given their search criteria. However the field is still in its early days, Johnston-Glenberg (2018) points out that the use of IVR in education is so new that design guidelines for implementing IVR it have not been published.

2.2.2 Characteristics and Features

IVR is commonly implemented through *Head Mounted Displays* (HMDs) where VR glasses/googles encase the user's field of vision within the virtual environment (often with the addition of peripheral devices such as headphones or controllers), and *Cave Automatic Virtual Environments* (CAVEs), where the device projects the image of the virtual environment onto the surfaces of the room containing the user. The latter is expensive, often immobile and requires a dedicated space, and as such it is far less commonly found in education (Freina & Ott, 2015; Johnson-Glenberg, 2018). IVR is a technology that comes with unique combinations of characteristics and resultant features not presently available through alternative means, the literature identifies many of them (Johnson-Glenberg, 2018; Mikropoulos & Natsis, 2011). Mikropoulos & Natsis (2011) identified several notable features, which are creating presence, first-order experiences, natural semantics, size, transduction, reification and autonomy. Presence is a sense of being in the VE, and is considered one of its most notable features (further explained in 2.5.2). First-order experience describes the opportunity for users to navigate the VE from a first-person point of view. Natural semantics describes the effect of representing concepts with more natural symbolism (like representing an atom by its shape) allowing users to avoid learning and remembering by the traditional approach of abstract symbolism. Size, transduction and reification relate to the capability of IVR to provide access to locations/experiences that would otherwise be unreachable, unfeasible, or impossible through ordinary means. Users can adjust the size of themselves or environments within a VE, allowing them access to microscopic or gigantic worlds, like an atomic nucleus. Transduction describes how VR can allows user to sense/feel data that would be beyond their experiences and sensory range in the real world, such as simulating a historic event or making the spectrum of non-visible radiation visible within a VE. Reification describes how abstract concepts and relations can be solidified as virtual objects, helping users construct conceptual relationships and visualise 3D situations (Lee, Sergueeva, Catangui, & Kandaurova, 2017). This can assists them in transitioning from representational to conceptual thinking within a domain (Hanson & Shelton, 2008). Depending on the characteristics of the simulation, autonomy describes how objects can be programmed to undergo interactions without the input of users, such as a planet rotating upon its axis in a solar system VE.

These features are made possible by certain characteristics of IVR technology. In summary, these are its capability to manage a high volume of data in a short time, present information in a dynamic-interactive format with multiple representations, and communication between users. There are other significant capabilities of IVR identified in the literature. Thanks to the interactivity (the feel, properties and quality of interaction) of IVR and the sense of presence it creates, it offers an opportunity for safe embodied learning within a limitless variety of environments (Johnson-Glenberg, 2018; Johnston et al., 2018; Markowitz et al., 2018). IVR can also represent the real world with physically impossible additions such cues or gamified components (Cooper et al., 2018; Kavanagh et al., 2017). Some of these characteristics are often the motivation to use IVR, including the access to locations/experiences (Freina & Ott, 2015), interactivity (Hanson & Shelton, 2008), collaboration, constructivism, gamification, presence, motivation & enjoyment, personalised learning and deeper learning (Kavanagh et al., 2017).

A notable format for IVR is *360˚ video*. As its name suggests, 360˚ videos are shot using an omni-directional camera and can be displayed through most devices so long as the point of view (POV) can be adjusted somehow. Notably, compatible hardware includes 2D monoscopic devices such as phones, where a user can adjust their POV by moving their device thanks to built-in rotational tracking (Feng, 2018). These videos can be combined with bidirectional sound to complete the experience. Of course, with the larger field of regard and head-based rendering HMDs are far more immersive than monoscopic devices for delivering these videos (Hendriks Vettehen, Wiltink, Huiskamp, Schaap, & Ketelaar, 2019). A significant advantage of 360˚ videos over graphically rendered, synthetic content is that their hardware and production costs are far smaller in comparison, and they are far easier to create (Feng, 2018; Kavanagh et al., 2017; Ulrich, Helms, Frandsen, & Rafn, 2019). They could be useful for content where visual realism is most important because, as captured video, they almost bypass this issue completely. There is also a need for low cost systems that incorporate as much school and student owned hardware as possible, something that 360˚ videos are better suited for as (Kavanagh et al., 2017). Relative to most 'full-scale' IVR, 360˚ videos are often less interactive (Ulrich et al., 2019), furthermore, there is a lack of educational research into them (Feng, 2018; Ulrich et al., 2019). Whether they count as IVR is also debatable, Kavanagh et al. (2017) consider that by the minimal definitions they do count, especially if viewed through an HMD.

2.2.3 Common Applications

Currently, IVR is applied for various purposes. Kavanagh et al., although reviewing VR in general, found that only 28% specifically related to general education, the rest in other domains including health, engineering and science (2017). As mentioned, VR has been tested in STEM education since the 90s (Markowitz et al., 2018), particularly astronomy and computer science, but is mostly limited to higher education. Training, the facilitation of the transfer of practical-psychomotor skills, is a common use for IVR (58% of identified simulations by Kavanagh et al.) as users can repeatedly train within a simulation of the real-world task without the physical and ethical risks. This is particularly important for medicine with the training of surgical procedures & techniques (Freina & Ott, 2015; Kyaw et al., 2019). For similar reasons IVR has been extensively used in military training. For instance, one simulation educated US soldiers on procedures pertaining to corrosion prevention and control (Webster, 2014), and of course flight simulators are another example (Lateef, 2010). As many affective skills require repetition to build, IVR has been employed in in areas related to social/psychological phenomena. These include treating irrational fears/phobias, PTSD, anxiety & personality disorders, and practicing stress management strategies (Jensen & Konradsen, 2018; Markowitz et al., 2018). Another application is simulated laboratories which have the potential to overcome some of the educational issues inherent with real labs and distance learning (Potkonjak et al., 2016). Museums have also been early adopters of IVR, using it for more interactive and immersive experiences (Kavanagh et al., 2017; Taxén & Naeve, 2002).

2.2.4 Issues Currently Faced by IVR

Although IVR devices have drastically reduced in price they are still expensive, especially when considering the limited budget and time of schools (Jensen & Konradsen, 2018; Johnston et al., 2018; Markowitz et al., 2018). A fact which is only exacerbated by training and overhead costs common to the technology (Kavanagh et al., 2017). Other notable issues include: training requirements; the complexity of development (Hanson & Shelton, 2008); issues related to specialised input hardware; recognition inaccuracies; insufficient realism; software usability; usefulness (with respect to learning outcomes); lack of engagement; ineffectiveness (Kavanagh et al., 2017). Another major consideration is *Cybersickness* which “can be described as a set of symptoms such as fatigue, headache, eye strain, stomach awareness or nausea that can occur during and/or after exposure to a VE” (Melo, Vasconcelos-Raposo, & Bessa, 2018, p. 160). Educators are required to mitigate risks of harm to students, as outlined by duty of care policy. Thus, cybersickness should be considered a serious issue when employing IVR for primary and secondary education, especially when using HMDs as they have been associated with the most severe cybersickness effects (Jensen & Konradsen, 2018; Melo et al., 2018). It is for this reason that many IVR products state their recommended age ranges, typically 12 or 13 years and up. Research into safety in IVR is ‘nascent at best’ and cybersickness is not the only issue, research has warned against exposing children to emotionally stimulating VEs and a risk of children misidentifying IVR experiences with memories of events in the real world (Southgate et al., 2019).

Many of these issues are clearly impermanent. Hardware prices will continue to drop and eventually IVR will save money in certain contexts (i.e. removing travel costs). The hardware and software will continually improve with research in design and developments in technology, mitigating issues such as poor usability and insufficient realism (Kavanagh et al., 2017). The issues of usefulness, engagement and ineffectiveness are arguably the product of a largely pedagogically uninformed development community. A common message in the literature is that very few applications and devices are grounded in pedagogy and educational theories, even if explicitly designed for education (Jensen & Konradsen, 2018; Johnson-Glenberg, 2018; Johnston et al., 2018; Kavanagh et al., 2017; Lu, Wu, Cheng, & Lou, 2018; Makransky et al., 2019; Parong & Mayer, 2018). Many applications take their cues from the work done in the VR entertainment industry, where maximising engagement is the most pressing objective (Johnson-Glenberg, 2018). As such many developers take a technology-centred approach rather than a student-centred one (Makransky et al., 2019). Johnston et al. summarises the present needs in using IVR for education: “To take fullest advantage of VR, educational leaders need to understand the pedagogical aspects of VR applications to distinguish and support different strategies and optimize learning for students.” (2018).

2.3 Working Memory

Working Memory (WM) is an important concept that is central to the research question, as it is a key component of later developments of CLT and related theories. In his primer ‘Working Memory’, Alan Baddeley summarises: “Working memory refers to the system or systems that are assumed to be necessary in order to keep things in mind while performing complex tasks such as reasoning, comprehension and learning” (2010, p. 136). WM acts as an interface between the external environment and information stored in *Long-Term Memory* (LTM), it is severely limited in its information storage capacity and duration. Many of the content-presentation implications come from the limitations and structure of WM.

2.3.1 2.3.1 History

‘Working Memory’ was originally coined in 1960 in the book ‘Plans and the Structure of Behaviour’ by Miller, Galanter and Pibram. During the 60s developments in computing inspired information-processing approaches to psychology, developing the field of cognitive psychology. WM was originally an extension of *Short-Term Memory* (STM) and the idea that information from the environment, having passed through sensory buffers, goes through a succession of storage systems starting with STM and ending in LTM, underlying the creation of memories and knowledge. Atkinson & Shiffrin (1968) proposed that STM acts as a WM which further controls the flow of information and acts as a short term store. However, the model had its limitations. It implied only a maintenance of information in STM which would ensure LTM storage regardless of the nature of the processing. In reality the degree of conceptual learning is greatly affected by the way it is processed. Furthermore, the model was insufficient when explaining the abilities of people with specific memory impairments. These issues could only be ameliorated if the STM somehow organises and processes the information it receives (Baddeley, 2010).

2.3.2 The Multicomponent Model

Through empirically driven developments, Baddeley & Hitch proposed the *Multicomponent Model* in 1974 to provide a broad theoretical framework for WM, Figure 1 depicts the updated model.

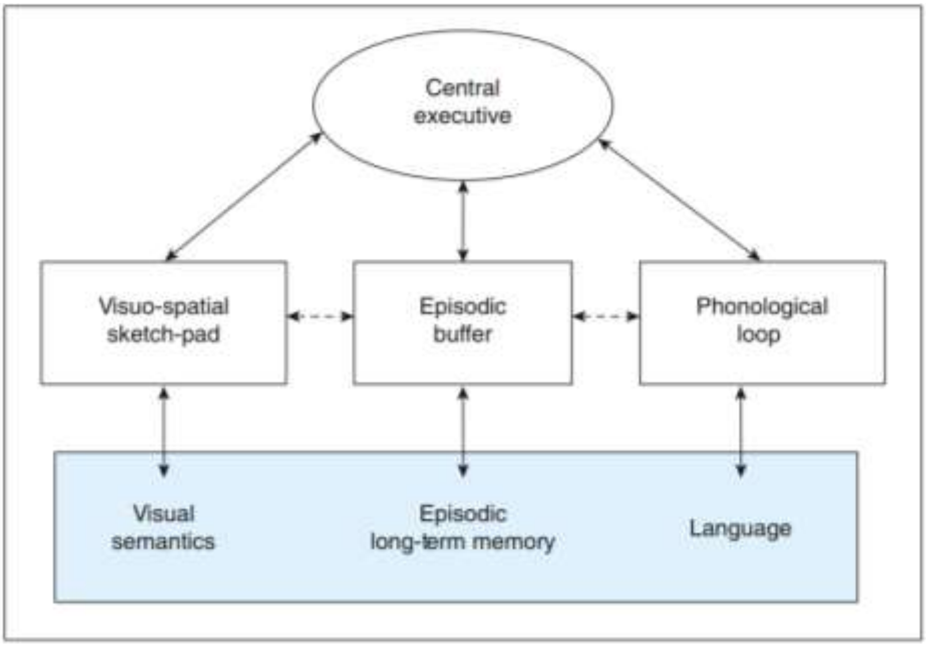

Figure 1: Multicomponent model

(Baddeley, 2010)

According to this model of WM, sound elements (especially language) are held in the *Phonological Loop* that continuously rehearses the sound to prevent it from decaying. The *Visuo-Spatial sketchpad* stores visual information elements and their features. This is known as the *Dual-Channel* theory which itself follows *Dual-Coding* theory (Paivio & Lambert, 1981), which describes that verbal and non-verbal information is processed differently and in semi-separate channels. The *Central Executive* is the least understood but most important component, it supervises and coordinates the other subsystems and is thought to be divided into a habitual side and an attentionally limited controller (Baddeley, 2003). The first significant difference between Atkinson & Shiffrin's model and the multicomponent model is that the WM is split between at least three subsystems with their own capacities, that also work in parallel rather than succession. To account for limitations in the model, the *Episodic Buffer* was added which can temporarily hold chunked information from LTM called *Episodes*. The concept of episodes comes from Tulving's work (1989) on episodic memory, episodes integrate information across both space and time, within WM information processed between the components is integrated into these episodes. The buffer is assumed to retrieve from and feed into episodic LTM, therefore through the medium of conscious awareness the central executive can access the episodic buffer to interact with information stored in LTM. The buffer also serves as an interface between the subsystems, as the two channels are thought to use their own coding system the buffer is thought to use a multidimensional code which can incorporate and integrate the two.

2.3.3 Implications & Limitations

The implications of the multicomponent model are that WM can create, manipulate and update LTM with new information from the senses (Baddeley, 2000), ultimately connecting the outside world and LTM. The emphasis of the model is that WM does not just activate old memories, but rather temporally retrieves and manipulates them. A significant drawback to WM as an information processing system is its severely limited storage. Each subsystem is limited in capacity and duration, and together are responsible for the limited and temporary nature of WM. Miller suggested that it can store 7 elements simultaneously (1956), later research refined the number to 7±2 (Paivio, 1986). Today, the phonological loop is thought to hold roughly 7 sound elements and the visuo-spatial sketchpad can hold roughly 3 objects & their features, depending on the complexity of the content in both cases (Baddeley, 2003). Due to the computational demands of its functions the episodic buffer is assumed to be limited to about 4 episodes (Baddeley, 2010). These capacities are thought to vary slightly between people based of measurements. If too much information is passed through the WM, either attention simply cannot be placed on it all or portions of that information will decay before processing and integration is completed, this will impair learning.

There are some theoretical disputes with the multicomponent model. The nature of the central executive is one of these disputations partly because it is the most abstract and the least supported by direct empirical evidence. Sweller et al. (2011) argue that the concept of an independent central executive leads to an infinite regress of central executives, to avoid this issue they argue that knowledge held in LTM acts as a de facto central executive. In any case the central executive will perform the same function, important for this study is that there is considerable evidence in support of the other subsystems (Sweller et al., 2011). WM also seems to deal with sensory memories and LTM episodic memories differently, the multicomponent model does not necessarily account for this difference. This lead Ericsson and Kintsch (1995) to postulate a new processor (long-term working memory) for when WM deals with information from LTM. However, given this studies focus on CLT it makes no difference whether WM is divided in such a way (Sweller et al., 2011).

2.4 Cognitive Load Theory

CLT is a cognitive/learning theory that describes how the human brain learns and stores knowledge (Westby, 2018). CLT is built on the assumptions of cognitive architecture, namely that the human brain can only process a limited quantity of information at one time, and that LTM can practically store a limitless quantity of information. The critical implications of CLT is that it can describe why some content presentation styles work and others don't, primarily based on the implicit complexity of information, how it is presented, and the prior knowledge of the learners involved. The principles of CLT as applied to teaching are supported by a significant body of research across many applications (Westby, 2018). CLT is central to this study as it will be the lens through which the platform will be analysed, providing a empirically supported theoretical basis to support the analysis and conclusion drawn.

2.4.1 History

CLT developed from research on problem solving by John Sweller (1988), who argued that instructional design can be used to reduce the cognitive load in learners. Sweller noticed that learners used means-ends strategies for solving problems, a strategy that not only used a lot of processing capacity but was not effective in developing knowledge. He suggested that instructional material should avoid unnecessary cognitive load, and support more efficient instructional styles such worked-examples, where learners could pick-up problem-solving methods from the instructor rather than using inefficient means-end approaches. Over the next 25 years, researchers developed a range of content presentation methods based off CLT and built on extensive research (Sweller et al., 2011). Other closely related theories were developed for more specific practices like multimedia learning (Mayer & Moreno, 2003) and mediated messaging (Lang, 2006). During the same time, methods to measure cognitive load were developed including indirect, subjective, and efficiency measures as well as secondary task performance and the more objective physiological measures.

2.4.2 Theory & Basis

CLT is based on a theory of cognitive architecture derived from evolutionary psychology (Sweller et al., 2011). It is mostly concerned with domain-specific skills that humans haven't necessarily evolved to learn seamlessly, or what Geary (2008) termed as *Biologically Secondary Knowledge*. Such knowledge is what educational institutions are concerned with, and often requires substantial cognitive effort to learn. Secondary knowledge is processed through a *Natural Information Processing System* which comprises several principles (Sweller et al., 2011, see part 2 for further details). Two principles are central to CLT. The *Narrow Limits of Change Principle* explains why cognitive systems need to have a limitation in the quantity of information they can pay attention to, which is the avoidance of combinatorial thought explosions. In cognitive psychology, this principle is realised with the limitations of WM, and is why WM is so important for CLT. *Cognitive Load* describes the cognitive resources, or WM resources, necessary to complete a learning task, CLT is centred on ensuring that the WM processes the total cognitive load of a task successfully (without exceeding its capacity). The *Environmental Organising and Linking Principle* describes how information stored in LTM is used to ensure that cognitive activity is appropriately coordinated with the environment. Once again, the WM provides the structure for this principle. Stored information, unlike information from the environment, is organised and does not need to be subjected to the same limitations. Information stored in LTM is structured in *Schemas*, described by *Schema Theory* (Sweller et al., 2011; Westby, 2018).

Westby summarises the nature of schemas: "A schema organizes elements of information, representing knowledge about concepts, objects, and the relationships they have with other objects, situations, events, sequences of events, actions, and sequences of actions." (2018, p. 9). By definition, schema are more complex configurations than episodes, CLT extends WM from recalling episodes to recalling schema. Learning, expertise and skill are due to the acquisition and development of schemas. This is because schemas contain a collection of information elements and their connections developed from previous learning (so the they don't have to be processed again). They can be recalled as a single element in WM thanks to the lack of necessary limitations for stored memories. Schema recollection and application is increasingly automated through practice, reducing the cognitive burden of this process (Josephsen, 2015; Westby, 2018).

Every well-established model of WM is based on separate visual and audial channels, including Baddeley's multicomponent model, as already discussed. As the nature of WM is central for CLT, cognitive load can be distributed between these two channels (Sweller et al., 2011, p. 130). Indeed, the ability to process visual and audial information simultaneously is likely a biologically primary skill, something we do effortlessly. This is known as the *Dual Modality*. Engaging both channels could effectively increase WM capacity, insofar that information is split between audio and visual modes. Both channels will still maintain their individual limits, however the process of integrating information from both (presumably within the episodic buffer) incurs a cognitive load cost (Mayer, 2009).

2.4.3 Categories of Cognitive Load & Resources

According to CLT, there are two types of cognitive load which are additive, and two types of resources allocated for processing (Sweller et al., 2011). *Intrinsic Load* relates to the inherent complexity of a learning task, information elements and their connections that are relevant to the learning task. *Extraneous Load* relates to the way the learning task is delivered to the learner, and largely describes the information processing that is unnecessary for learning. The levels of both are influenced by *Element Interactivity*: “Interacting elements are defined as elements that must be processed simultaneously in working memory because they are logically related” (Sweller et al., 2011, p. 58), elements are ‘characteristically’ schemas. Low element interactivity tasks consist of elements that can be understood on their own, even if there are many elements they can be processed separately in WM with enough time. In contrast, high element interactivity tasks require learners to process elements that don’t make sense in isolation, taking up more space in WM to process them and their interactions simultaneously. Element interactivity provides an explanation of the capability of schema to reduce cognitive load: “Prior to a schema being acquired, those sub-elements must be treated as individual elements in working memory. After they have been incorporated into a schema, that schema can be treated as a single element in working memory” (Sweller et al., 2011, p. 58). A schema can ‘package’ a set of elements and interactions, creating a new individual element from their combination and effectively lowering element interactivity of a task when recalled to WM. Intrinsic load is increased if element interactivity is high, which characterises complexity, and extraneous load can be increased if the instructional approach or learning task has a higher element interactivity than necessary. This includes inefficient instructional approaches and task irrelevant information such as distracting, misleading or redundant elements. The mind will not automatically distinguish between task relevant and irrelevant information, not without conscious effort. Thus, the extraneous load imposed on learners is greatly influenced by the design of a learning task.

Germane Resources, sometimes called germane load (perhaps inappropriately), describes the cognitive resources allocated for intrinsic load processing, that is, it is the resources spent relating relevant information elements from LTM to new elements in WM and creating/updating schemas (Leppink, van Gog, Paas, & Sweller, 2015). Many studies consider germane Load as a third additive load source, however Kalyuga (2011) argued that it is redundant given the definition of intrinsic load and is more accurately considered a resource (P. Kirschner, Sweller, Kirschner, & Zambrano R, 2018; Sweller et al., 2011). The amount of germane resources allocated will also depend on the motivation, attitudes and characteristics of a learner (Kalyuga, 2011). Naturally, *Extraneous Resources* defines the resources allocated for extraneous load processing. The less resources spent on extraneous processes, the more resources available for germane processing and thus the WM can process a larger intrinsic load. These resources exist in a different dimension to the load types, but are considered together within a dual framework. Figure 2 represents the additive nature of intrinsic and extraneous load, as well as overload.

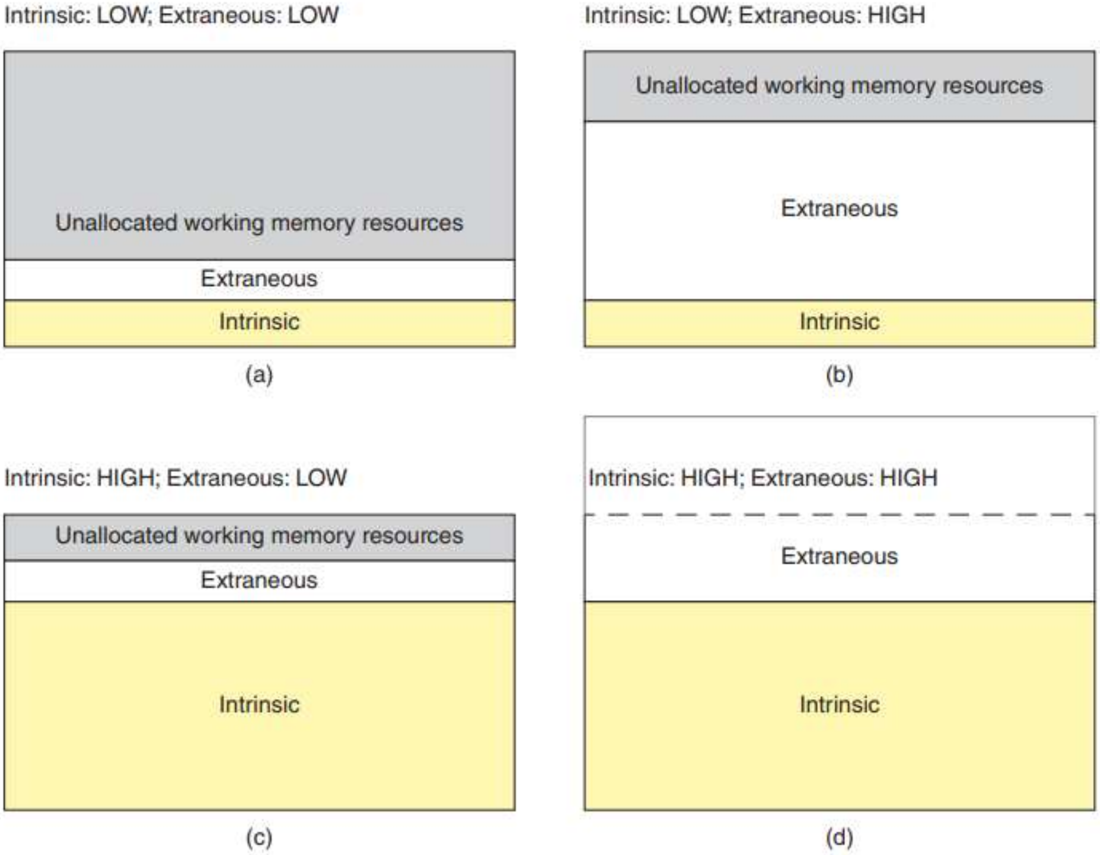

Figure 2: Intrinsic and extraneous cognitive loads as two additive types of cognitive load

(Leppink et al., 2015)

2.4.4 Instructional Implications of Cognitive Load Theory

If the cognitive load required for a learning task exceeds the limits of WM, learning will be impaired or fail completely (Sweller et al., 2011), this is often known as *Cognitive Overload*. Clearly, the avoidance of overload is a principle concern for teachers. CLT has contributed three main recommendations for instructional design: "present material that aligns with the prior knowledge of the learner (intrinsic load), avoid nonessential and confusing information (extraneous load), and stimulate processes that lead to conceptually rich and deep knowledge (germane load)" (de Jong, 2010, p. 111).

High intrinsic load can overload learners. As schema acquisition allows learners to reduce the element interactivity of a task, students who lack background schema may encounter such a high degree of element interactivity that the intrinsic load of the task will exceed their WM limits, regardless of design of the learning task. This can be avoided if the students' prior knowledge is considered so their learning capabilities can be validly estimated, otherwise intrinsic load cannot be altered without changing the learning objectives of the task itself. After the intrinsic load of a task is altered to fit the prior knowledge of the students, the main objective for teaching is to reduce extraneous load such that maximal germane resources are committed to intrinsic load processing. This is especially important as teachers have primacy of influence on content presentation in learning tasks (Sweller et al., 2011). Numerous methods have been identified for altering intrinsic cognitive load and promoting germane processing, but for the purposes of this study, reducing extraneous load in the presentation of content will be the focus. A series of basic cognitive load effects that determine extraneous load have been identified and empirically supported, they are summarised in Table 1.

Table 1: Traditional effects studied by CLT and why they reduce extraneous cognitive load

Effect	**Description**	**Extraneous Load**
Split-Attention Effect	Replace multiple sources of information (frequently pictures and accompanying text) with a single, integrated source of information	Reduces extraneous cognitive load because there no need to mentally integrate information sources and or apply search strategies
Modality Effect	Replace a written explanatory text with another source of visual information such as a diagram (unimodal) with a spoken explanatory text and a visual source of information (multimodal)	Reduces extraneous cognitive load because multimodal presentation uses both the visual and auditory processors of working memory

Redundancy Effect	Replace multiple sources of information that are self-contained (i.e. they can be understood on their own) with one source of information	Reduces extraneous cognitive load caused by unnecessarily processing redundant information

(Van Merriënboer & Ayres, 2007)

It is important to understand that these effects will only occur under the right conditions. Identifying sources of extraneous load is not a simple task either, especially if the prior knowledge of the learners is not properly accounted for. For instance, various researchers tried to replicate the findings of Sweller and colleagues without success. Although this was partly due to inherent differences in the methods applied (such as the prior knowledge of participants), and because the assumed levels of cognitive load were also inaccurate. That is, the modified tasks that were thought to have lower extraneous load and correctly follow the recommendations of CLT actually, upon closer inspection, produced more (May, 2004).

2.4.5 Cognitive Load Theory & Collaboration

Kirshner and colleagues (2018) used CLT to provide a theoretical basis to collaboration. According to the principle of *Mutual Cognitive Interdependent* collaboration creates a *Collective Working Memory* (F. Kirschner, Paas, & Kirschner, 2011). This collective WM is created through the communication and coordination of the group members, who can distribute the elements and interactions of a task between themselves. However, the process of communication (transaction) can create extraneous load. The effectiveness of communication greatly influences the success of collaboration. Kirshner et al. named 9 collaborative cognitive load principles. For this study these principles were kept in mind for the collaborative capabilities of the evaluated platfrom.

2.4.6 Cognitive Theory of Multimedia Learning

Mayer & Moreno (2003) proposed the *Cognitive Theory of Multimedia Learning* (CTML) to assess multimedia educational material. CTML is based off the dual channel and dual modality assumptions, that the channels are limited in capacity, and that meaningful learning requires substantial processing. According to Mayer & Moreno, *Multimedia Learning* is learning from words and pictures. The theory is summarised by Figure 3. CTML is based on the same assumptions as CLT (Kalyuga, 2011) and is focused on cognitive load and overload, it suggests that there are three types of cognitive processing which are *Essential*, *Extraneous* and *Generative* processing. These correspond to the three types of load traditional used by CLT (intrinsic, extraneous, germane). Generative processing is also dependent on learner related and affective factors (Mayer, 2014), like germane resources. However, Kalyuga argues that intrinsic load is related to both essential and generative processing, reducing from 3 to 2 categories of processing with the addition of germane resources allocated to learning-related processing. Responding to a direct email inquiry regarding CTML and CLT, Mayer stated that he and Sweller ‘basically agree’ although they use somewhat different terms (May, 2004). CTML defines principles for increasing processing efficiency that are essentially identical to the effects described by CLT, but are specifically focused on multimedia learning. The number of principles was expanded to 12 with each supported by empirical evidence (Mayer, 2009). Those that are related to reducing extraneous load in the presentation of content, and associated methods, are summarised in Table 2.

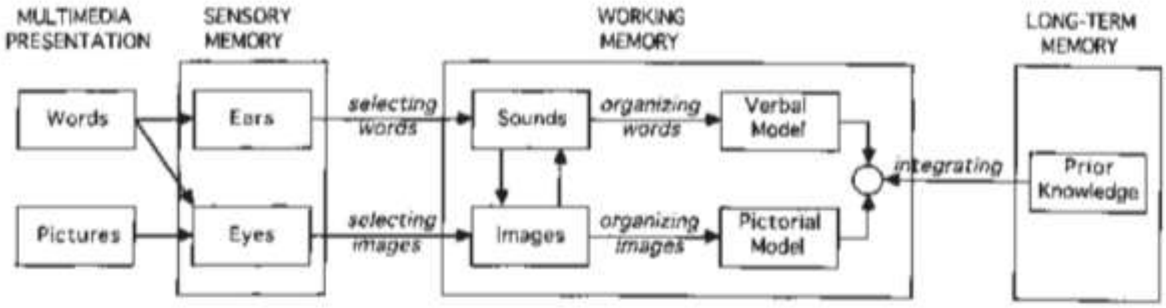

Figure 3: The Cognitive Theory of Multimedia Learning

Table 2: Principles studied by CTML and example methods

Effect	Description	Content Presentation Methods
Coherence Principle	People learn better when extraneous material is excluded rather than included - Redundancy	Weeding: Eliminate interesting but extraneous material to reduce processing of extraneous material: - Minimise necessary words, images, sounds and details - Consider removing task irrelevant words, images, sounds and details
Signaling Principle	People learn better when cues that highlight the organization of the essential material are added - Split Attention	Signaling: Particularly when weeding isn’t feasible, provide cues for how to process the material to reduce processing of extraneous material:

	- Modality	- Stress key words in speech - Organizing text with outlines and headings - Visually distinguish important information - Typographical cue - Orienting cue - Temporal cue - Spotlight - Label - Flash/Color change text - Zoom - Cue as feedback - Providing cues in multiple modalities
Redundancy Principle	People learn better from graphics and spoken text (narration) than from graphics, spoken text, and written text - Modality & Redundancy & Split Attention	Eliminating redundancy: Avoid presenting identical streams of written and spoken texts
Spatial Contiguity Principle	People learn better when corresponding words and pictures are placed near each other rather than far from each other on the page or screen - Split Attention	Aligning: Place printed words near corresponding parts of graphics to reduce need for visual scanning: - Pop-up text
Temporal Contiguity Principle	People learn better when corresponding words and pictures are presented at the same time rather than in succession - Split Attention	Synchronizing: Present spoken text and corresponding animation simultaneously to minimize need to hold representations in memory
Segmenting Principle	People learn better when a multimedia lesson is presented in user-paced segments rather than as a continuous unit	Segmenting: Allow time between successive bite-size segments: - Have a break point

	- Split Attention & Modality	- Allow users to re-watch earlier content
Modality Principle	People learn better from graphics and spoken text than from graphics and written text - Modality & Split Attention	Off-loading: Move some essential processing from visual channel to auditory channel
Multimedia Voice Principle	People learn better from words and pictures than from words alone - Modality	
Image Principle	People do not necessarily learn more deeply from a multimedia presentation when the speaker's image is on the screen rather than not on the screen - Redundancy	

(Ibrahim, 2012; Mayer & Moreno, 2003; Sorden, 2012; Sweller et al., 2011; Wouters, Paas, & van Merrienboer, 2008)

2.5 IVR & Learning

As discussed in 2.2.4, although there has been a lot of research into IVR there is a deficiency of education-centred design and research that is grounded in pedagogy. A further issue is the quality of research, in a review of the literature Jensen & Konradsen (2018) found that, for their sample, studies were lower than average in quality based off their metric, primarily because many of the studies relied on user evaluations and the validity of measurement instruments used. In any case, for learning IVR is assessed by the concepts of *Immersion* and *Presence* (Cooper et al., 2018; Jensen & Konradsen, 2018; Johnson-Glenberg, 2018; Mikropoulos & Natsis, 2011), this is the same presence described in 2.2.2. As Cooper et al. explain, "The principal aim in designing VR systems is to immerse users to such an extent in the virtual worlds that they accept the virtual world as 'real'." The terms immersion and presence are often debated with some conceptual overlap between definitions, sometimes they are used almost interchangeably. The following paragraphs describe the most rigorous definitions of these terms and the theoretical reasons why they might benefit educational outcomes. As a note, for this review, learning is divided between cognitive (Anderson et al., 2001), affective (Krathwohl, Bloom, & Masia, 1964) and psychomotor (Harrow, 1972).

2.5.1 Immersion

Slater defines immersion as “the objective level of sensory fidelity a VR system provides” (Cooper et al., 2018). By this definition immersion is a quantifiable description of VR and is influenced by both the match between sensory data with a user’s subjective world models, and the match between proprioception & sensory data (Slater, Linakis, Usoh, & Kooper, 1999). This technical definition is often adopted (Johnson-Glenberg, 2018; Krokos, Plaisant, & Varshney, 2018; Kyaw et al., 2019; Markowitz et al., 2018; Menin, Torchelsen, & Nedel, 2018; Tcha-Tokey, Christmann, Loup-Escande, Loup, & Richir, 2018). In contrast, Witmer & Singer define immersion as a “psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences.” (1998, p. 227). These two definitions differ in that Slater’s is technical whereas Witmer & Singer’s is phenomenological, but the connection between the two is quite clear (Johnson-Glenberg, 2018). From Slater’s definition, factors that influence immersion are dependent on technological features such as the field of regard, field of view (FOV), stereoscopy, display size, display resolution, head-based rendering, frame rate, refresh rate, visual realism and the realism of interactivity (Cooper et al., 2018). According to Witmer & Singer, factors that influence immersion include isolation from physical environment, perception of self-inclusion in VE, natural modes of interaction and control, and the degree of self-movement. Interaction & control can include features like the ability to stop, start, replay and manipulate visuals at a chosen pace (Thompson, Wang, Roy, & Klopfer, 2018). Both definitions describe a spectrum of immersion, as such Muhanna classified three ‘levels’ of immersion (2015). “The lowest level of immersion is in basic virtual systems, such as hand- and monitor-based systems. Partially immersive systems use a single projector to display a virtual world on a large screen. Fully immersive systems provide 3D virtual scenes in a large field of view.” (Menin et al., 2018, p. 58). Any 3D scene displayed through an HMD can be considered fully immersive by this definition.

2.5.2 Presence

Presence is defined as “the subjective experience of being in one place or environment, even when one is physically situated in another.” (Witmer & Singer, 1998, p. 225). The power of presence is explained by Markowitz: “Together, an immersive virtual experience with high levels of presence allows the individual to suspend any belief that the experience is mediated.” (2018). According to Witmer & Singer: “Presence in a VE depends on one’s attention shifting from the physical environment to the VE, but does not require the total displacement of attention from the physical locale.” (1998, p. 226). This means that presence can come in different degrees, the more attention is focused on the VE the more presence a user will feel. That is, presence depends on the ability to focus on ‘one meaningful, coherent VE stimulus set’ and is influenced by a sense of involvement and immersion, which act as interdependent-necessary conditions (Witmer & Singer, 1998). Involvement relates to focus which is affected by any personal problems or activities occurring outside the VE, immersion clearly relates to the nature of the stimulus set. Resultantly, presence relates to the allocation of attention and, therefore, cognitive resources (Cooper et al., 2018). The more present a user feels within an immersive experience, the more cognitive resources they will allocate to process it. Empirical research has found positive correlations between presence, involvement and immersion (Cooper et al., 2018; Moreno & Mayer, 2002; Tcha-Tokey et al., 2018). Jensen & Konradsen found that presence was also impacted by an awareness of people watching, sitting as opposed to standing, and is reduced in more anxious or reserved individuals. Using IVR often creates a sense of presence instantaneously, for this reason presence is often considered the most important feature for IVR, as most IVR is highly immersive (Johnson-Glenberg, 2018).

2.5.3 IVR Pedagogy & Affordances

According to several learning theories, presence and immersion should lead to improved learning outcomes, this comes down to the capabilities of IVR to fulfil their requirements and recommendations. The theories most commonly applied when creating and designing IVR, even if only tacitly, are experiential learning, discovery learning, constructivism, and situated cognition (Johnston et al., 2018). Each of these learning theories encourage active, authentic, reflective and interactive learning experiences. Since it is thought that users might perceive an immersive environment as unmediated they will respond to stimuli as they would in the real world (Lee et al., 2017; Markowitz et al., 2018), IVR can provide the types of learning experiences these theories encourage. The role of immersion and presence is that they enhance the authenticity and interactivity of these experiences, the more immersive a VE the more realistic and or authentic it will be, the more present a user feels the more they will treat the IVR experience as real. Furthermore, IVR’s features of size, reification and transduction extend the possible experiences beyond what can be done in the real world. Mikropoulos & Natsis (2011) identified constructivism as the most commonly applied of these theories for the design of educational VEs. Constructivism is closely related to experiential and discovery learning, but with the construction of personal meaning as an additional step. In constructivism, learners are actively involved in knowledge construction by organizing input from experiences which they can then apply in different contexts and environments. They identified that five affordances of VR, including spatial knowledge representation, experiential learning, engagement, contextual learning and collaborative learning that closely follow the recommendations of constructivism.

IVR should produce superior affective outcomes as studies suggest that people internalize their virtual interactions as real (Markowitz et al., 2018). Highly immersive experiences can induce *Transportation*, transported users lose track of time, lose access to real-world facts, and experience strong emotions (Feng, 2018). The factors that created presence also create *Arousal*, a state of mind associated with more affect driven and heuristic thinking and less thoughtful information processing (Hendriks Vettehen et al., 2019). According to the theory of *Optimal Arousal*, users are more likely to enjoy media content if they feel present (Hendriks Vettehen et al., 2019). According to the *Expectancy Model* (Pintrich, 2003) this increased enjoyment, the novelty of IVR, the ease at which IVR content can be made interesting, and its interactivity can increase motivation. Motivated students tend to learn better (Kavanagh et al., 2017), put more effort into understanding and are more resilient to obstacles, which cause the allocation of more cognitive resources to a task. Modest positive correlations between self-reported motivation and grades have been found (Parong & Mayer, 2018).

IVR finds support in *Embodied Cognition*: "people can learn, form more positive associations toward a stimulus, and internalize information by acting in a space and performing motions that are relevant to a specific context relative to simply internalizing information about a phenomenon." (Markowitz et al., 2018). Embodiment stresses that the mind and body are 'inextricably linked'. When learning signals are linked with a motoric modality more neural pathways may be activated, resulting in improved learning (Johnson-Glenberg, 2018), a plethora of neuroscientific evidence supports embodiment. For the same reasons, embodiment and presence within locations can improve memory retention and recall (such as memory palaces), a phenomenon that can be achieved in a sufficiently immersive VE (Krokos et al., 2018). IVR is fully capable of creating experiences that include embodied learning.

2.5.4 Impact of IVR on Learning

Studies that investigate the impact of IVR on educational outcomes have produced a range of results. This is partly due to how new the area is, the spectrum of research topics, the range of measurement instruments, and variation in the nomenclature, to name a few. Despite this, there are noticeable trends in the findings. IVR generally produces positive outcomes in the spatial & visual aspects of cognitive learning (Jensen & Konradsen, 2018; Menin et al., 2018), psychomotor skills (Kyaw et al., 2019; Menin et al., 2018), and is powerful for producing positive affective outcomes such as motivation, interest, enjoyment, confidence and enthusiasm (Jensen & Konradsen, 2018; Madrigal, Prajapati, & Hernandez-Prera, 2016; Menin et al., 2018). The efficacy of IVR in the affective dimension seems to depend more on the simulations ability to evoke a response than interactivity (Jensen & Konradsen, 2018). The results of IVR on the knowledge and understanding component of cognitive learning outcomes is mixed. Jensen & Konradsen found that all the studies in their sample focused on skills that are lower level in Bloom's taxonomy and concluded that the IVR experience can often overshadow knowledge acquisition. Roussou (2009) used IVR with a CAVE system to teach abstract mathematical concepts and problem solving. No quantitative differences were found between interactive-VR, passive-VR (with a virtual guide) and non-VR, however qualitative differences were observed. In the interactive VR students made decisions intuitively supported by the environmental cues and feedback, allowing the students to use VR representations over the normal symbolic language of maths. However, no evidence suggested interaction lead to deeper learning. Markowitz et al. (2018) found that IVR created positive knowledge outcomes and that interactivity increased the demand for more conscious and deliberate processing, but their study did not compare IVR to other instructional methods.

Research, across various fields, has been conducted on 360˚ video. Lee et al. (2017) found that 360˚ (HMD) and regular video did not differ with respect to novelty, reliability and understandability of content, but was more enjoyable and interesting. Ulrich et al. (2019) found no difference between traditional teaching and 360˚ (HMD) & regular video in academic performance, students were most satisfied with traditional teaching and were equally satisfied with 360˚/regular videos. They concluded that this was likely due to a lack of interactivity with their 360˚ videos. Emotionally, 360˚ video and traditional teaching outperformed regular video. Yoganathan et al. (2018) compared 360˚ (HMD) and regular video for presenting a tutorial for surgical knot tying skills. Participants who viewed the 360˚ video had significantly better knot tying skills, which persisted when the videos were combined with face-to-face teaching. Feng (2018) compared a 360˚ narrative video ad (2D screen) with a regular narrative video ad. Specifically, transportation, ad usefulness, emotional responses, ad attitudes and brand attitudes were measured. The difference between the two video types was found to be significant when perceived ease of navigation was considered. The 360˚ video ad outperformed the normal video ad when viewers navigate smoothly, use visual and auditory cues, know where the desired point of view originates and have control over viewing direction. Hendriks Vettenhen et al. (2019) tested the effect of a 360˚ video news story (HMD) on presence, enjoyment, credibility, recognition, and understanding, relative to the same video in 2D. The 360˚ videos improved presence, enjoyment and credibility with no negative effects on recognition and understanding, enjoyment and credibility were found to be mediated by presence. The researchers conclude that understanding was not impacted because of a combination of the low information density of the video and a suppression effect of presence. It is worth mentioning that both video types were viewed through an HMD.

2.5.5 IVR & Cognitive Load

What is clear from the research is that IVR has the potential to improve spatial and procedural learning, immersion & presence, affective and psychomotor outcomes. However, studies so far have shown that IVR usually does not outperform traditional methods and non-IVR with respect to the cognitive outcomes of knowledge and understanding. This is likely because IVR can be overstimulating and distracting as a realistic, 3D, 360˚ experience, will deliver substantially more information to learners than other instructional formats (Jensen & Konradsen, 2018). Given that CLT is concerned with the transfer of knowledge from WM to LTM, especially the presentation of content, it is natural to consider applying CLT, and theories based off it, to the design of educational IVR software.

A few articles directly applied CLT and CTML in their studies of IVR. Makransky et al. (2019) carefully compared students undertaking a virtual laboratory using non-IVR through a desktop pc and IVR through a HMD, measuring cognitive load directly with EEG. They found that students learned less through IVR and experienced significantly higher cognitive load, but experienced a greater sense of presence and enjoyment. Furthermore, they found no redundancy effect when spoken-written text was used instead of just written text but concluded that this might be because students simply ignored the text during the spoken-written condition. Parong & Mayer (2018) compared IVR through HMDs and well-designed slide-shows, using CTML and the Expectancy Model to inform their hypothesis and analysis. IVR produced significantly higher motivation, interest and engagement but scored significantly worse in the post-test, especially for factual questions, most likely because of high extraneous load and less learner control. Lu et al. (2018) measured the impact of varied learning-support design strategies on the efficacy of a middle-school biology IVR. They found that equipment operation, learning strategies and high visual fidelity increase cognitive load. Moreno & Mayer (2002) studied the effect of spoken and on-screen text explanations presented through either desktop or HMD. They found that HMDs produced greater presence but retention, transfer and program rating were unaffected by display type and therefore the level of immersion. A modality effect was found, students scored higher on retention when spoken text was used instead of written text or both. Like the previous research, these studies show that IVR does not necessarily improve knowledge and understanding, being outperformed by non-IVR methods in these outcomes. Furthermore, cognitive load was significantly increased in the IVR condition, as expected. The reasons why IVR tends to increase cognitive load relate back to the effects described in 2.4, specifically the split-attention, modality and redundancy effects.

The levels of immersion and presence can affect extraneous load. Makransky et al. suggest that the effect different levels of immersion have of cognitive load and learning outcomes requires more research. Numerous studies claim that a lack of realism/fidelity, which increases immersion, is an important problem for IVR (Kavanagh et al., 2017). However greater realism will increase extraneous load by increasing the levels of redundant, irrelevant and distracting elements. When cognitive outcomes like knowledge transfer are most important increasing realism could be detrimental (Lu et al., 2018; Moreno & Mayer, 2002; Parong & Mayer, 2018). Immersion can act as a seductive detail that creates a 'bells and whistles' novelty effect (Makransky et al., 2019), which can affect information processing in the same way that redundant text can. Similarly, Cooper et al. found numerous studies that suggest realism may not be as important for overall performance, and that reducing realism can lowers a VE's complexity, enhancing focus on the learning relevant information. The motivating effect of presence can increase the allocation of cognitive resources to germane processing, however increasing presence requires increasing immersion and thus extraneous load (Moreno & Mayer, 2002; Whitelock, Romano, Jelfs, & Brna, 2000). Makransky et al. based their study on this notion and found that, for an IVR science lab, the effect of increased extraneous processing outweighed the effect of the increased allocated cognitive resources, reducing learning despite producing affective benefits. Furthermore, as features that increase presence tend to increase arousal, the cognitive resources needed for learning can be reduced as arousal causes the brain to engage in less thoughtful cognitive processing. Schrader & Bastiaens (2011) found that for immersive educational games presence positively affected more trivial learning outcomes, but its emotional effects seems to be insufficient for complex learning outcomes.

Interactivity is another important factor of immersion that may not necessarily improve learning in certain contexts. For instance, Roussou found that for learning abstract problem-solving, interaction was less powerful than the non-interactive and explicit guidance of a virtual instructor. The risk inherent with interaction is that it can consume cognitive resources that could be used for germane processing, even low interactivity environments can overload learners. For instance, for 360˚ ads, Feng found that visual navigation can take up cognitive resources that are also needed for critically processing the ad. A difficult to navigate ad can overload viewers, resulting in confusion and frustration which in turn compromises affective learning. This is because search strategies require cognitive resources, and navigating a visual space can split-attention for mutual dependent information elements. Baumeister et al. (2017) found that visual navigation might be an inherent issue for HMDs. They found that a limited FOV created extraneous load from having to search beyond the limits of the display's FOV, an issue that is less prevalent when FOV is maximised. A core aspect of the problem is the difficulty in determining whether the cognitive costs of interaction are from usability/intuitiveness or the nature of interactivity (Makransky et al., 2019), it is possible that as interactivity becomes more natural it becomes less mediated, reducing the cognitive costs.

However, reviewing VEs, Mikropoulos & Natsis (2011) state: “carefully designed learning activities are more important than an exotic interface that contributes to intuitive interaction”. In any case, there are examples of low interactivity VEs, like 360˚ videos, producing cognitive learning outcomes that are comparable or superior to highly interactive VEs (Madrigal et al., 2016).

CLT encourages many additions that might reduce the realism of a VE. Cues are particularly important as they can help users identify task relevant information (Xie et al., 2017), which can reduce search time and assist students in segmenting the presented information (W. Roberts, 2008, For an overview on the advantages and disadvanatages of cueing). Cooper et al. found that substituting in information bearing cues at the cost of decreased realism enhances informational content and learning outcomes, with a reduction in fidelity reducing extraneous load. However, no significant effect was found when visual cues were used alone, probably because of the simulation’s high visual load. Multimodal (visual-audial-tactile) and bimodal (visual-audial & visual-tactile & audial-tactile) cues produced significant positive effects, which is consistent with the modality effect of CLT. They also found that the subjective sense of presences was greatest when multimodal & bimodal cues/feedback was provided, demonstrating that realism can be less significant for creating presence than informational content. The effect of cues has been studied in the training of psychomotor skills, where research has shown that richer cues and multimodal feedback increase the likeliness that VR trained skills will transfer to the real world (Moreno & Mayer, 2002). Like cues, visually guiding the attention of users towards objects in a 360˚ environment is another way to cognitively improve a task, as the effort spent searching creates extraneous load. This may also reduce the realism of a VE by placing an unnatural information bearing device in the VE. Bork et al. (2018) describes and compares various efficient approaches to visual guidance that improved search speed and decreased cognitive load. As well as explicitly guiding users, the design of an environment can influence the efficiency of searching. Olk et al. (2018) measured visual search and distraction in IVR. They found that visual reaction time increased when the discriminability between the target object (the one users needed to find) and its distractors decreased, that is, visual search is minimised when target objects were more distinguishable from everything else. Again, simulator realism could be reduced by making target objects clearly distinct from the environment. An interesting finding by Whitelock et al. (2000) was that their students, when searching for species in a woodland simulation, wanted conceptual tools that would help their search that wouldn’t be available in real life, and were frustrated by the absence of such tools in the VE.

2.6 Summary

Although IVR is effective for various learning outcomes, evidence suggests that it might be outperformed by non-IVR and traditional methods other for the cognitive learning outcomes of knowledge and understanding. The reason for this is likely because IVR can greatly increase cognitive load, both intrinsic and extraneous, negatively affecting schema development. IVR is highly immersive, but this immersion comes with a substantial amount of visual information that can overload or distract users from task-relevant information. Furthermore, some of the beneficial effects of IVR, such as the sense of presence and improved affective outcomes, require cognitive processing that could be used developing schemas. These outcomes can themselves be prevented if users are overload. As Lu et al. summarise: “To sum up, how to maintain the sense of presence, reduce cognitive load, and enhance learning effectiveness in a VR is an issue that should be solved when VR is applied into teaching.” (2018, p. 1306). Although more research is needed to determine the conditions in which learning outcomes within an IVR are optimised, unless the objective of a task or instruction is purely novel, reducing cognitive load in IVR is still important. The traditional recommendations of CLT find support in the literature despite its limitations and youth. A modality effect has been demonstrated, and design features that are consistent with split attention & redundancy such as reducing the information density, cueing, visual guidance, and visually distinguishing target objects produced superior learning outcomes for IVR. Likewise, Parong & Mayer’s results imply that many of CTML’s recommendations apply to IVR. Of these methods, they focused on the coherence and segmenting principle for their study into IVR, however it can be argued that all of the effects listed in Table 2 relevant.

For these reasons, this study applied the principles of CLT and CTML to the design of IVR content presentation, in the effort to reduce extraneous cognitive load and improve knowledge and understanding.

2.6.1 Gap in the Literature & Research Questions

As of the time of this study, an attempt to apply the principles of CLT directly to the evaluation of an educational IVR platform had not been made. Therefore, this study was unique in undertaking an evaluation of this kind, with CLT as a theoretical basis and benchmark. The following research questions provide the basis for the evaluation:

Q1 How does the educational IVR platform compare to the recommendations of CLT regarding the reduction of extraneous load in the presentation of content?

Q2 Where & how could the educational IVR platform be changed to better meet the recommendations?

3 Methodology

The aim of this research was to critically analyse and evaluate a educational IVR platform by applying a set of criteria from CLT & CTML. Specifically, the instructional design of the content/scenes developed for it was interrogated against the criteria. By analysing the content, a general evaluation of the platform its characteristics was to be made. Because of the evaluative nature of the study, a document analysis approach was adopted as it most closely fit the needs of the study. In this case the 'document' was the platform. Typically document analysis is used to determine meaning within text and images (Bowen, 2009), for this study the method determined where the platform's content met the prescribed criteria based of a direct use of the platform. The criteria were synthesised and applied with the relevant characteristics of the platform in mind. To do this, a description of the platform was made with its inherent characteristics described. After, CLT was used to design the criteria, however few guidelines describe how to adapt learning theories and pedagogy for the best use of IVR (Ritz & Buss, 2016), including CLT. As such the criteria used for this research was synthesised by following the most relevant recommendations from the literature with respect to the identified characteristics. Table 3 contains a list of common terms and abbreviations, some already described.

Table 3: Common terms

Term	Meaning
Immersive Virtual Reality (IVR)	A VE that systematically maintains an illusion of presence, such that users feel their bodies are inside the virtual environment
Head Mounted Display (HMD)	A display device, typically immersive, worn on the head or as part of a helmet, that has a small display optic in front of one (monocular HMD) or both eyes (binocular HMD)
Field of View (FOV)	The extent of the observable world that is seen at any given moment. In the case of optical instruments or sensors it is a solid angle through which a detector is sensitive to electromagnetic radiation
Point of View (POV)	The direction and location in which the FOV is centred on.
Three Degrees of Freedom (3DoF)	A VR device that can track directional rotation in three dimensions thanks to inbuilt sensors. But, as opposed to 6DoF, a user cannot translate their position within the VE, they are 'fixed in place'

Learning Management System (LMS)	A software application for the administration, documentation, tracking, reporting, and delivery of educational courses, training programs, or learning and development programs
Tethered Device	HMDs that require a constant connection to a powerful computer, typically to access the computer's graphical hardware and sensors to achieve 6DoF in a high fidelity VE

3.1 Subject Platform

The educational IVR platform was designed to provide both a small number of users a highly immersive IVR experience, or a large groups of users a less immersive IVR experience simultaneously, while maintaining a instructor's control over the simulation. It can be used on a range of hardware, which will dictate the level of immersion accessible. The Samsung GearVR HMD was focused on for this study. GearVR is an untethered HMD that connects to a smart phone (compatible Samsung Galaxy devices), which acts as the HMD's display and processor. The HMD itself acts as the controller, via a touchpad, back button and home button on its side, and contains the FOV, rotational tracking, proximity sensor and wheel for adjusting focus. A group of these devices can simultaneously connect to a server via wireless router and the content is controlled by a single device, preferably an ordinary computer or tablet, by accessing the server and opening the LMS. The content supported by Platform is primarily dependent on the hardware used, devices with different levels of interactivity can access different content. GearVR restricts users to the content that can use the touchpad, that is, scenes where users can at most look at objects within the environment and interact with them by tapping the touchpad.

Once a user is connected, they choose from a selection of avatars and enter their username/ID. users then enter a waiting/landing/meeting room, that they share with other connected users where they can adjust to the controls and experience. Without peripheral devices GearVR supports 3DoF thanks to sensors within the phone and HMD. Translational movement (6DoF), or the POV following with the user's head movements, is not supported under these conditions, so students are fixed in place unless the VE moves around beneath them. Through the LMS, the instructor can deploy immersive content to all the connected headsets simultaneously, the instructor can watch the content in real time on the regular display of their device. When a scene is selected a loading screen is displayed to the HMDs, who are then transported to the content. Each user has a cursor fixed at their centre of vision which moves with their POV (as they rotate their head), users can move this cursor onto interactable objects (essentially by looking at one) and select one by tapping the touchpad. Users can also make a waving gesture with their avatar by sliding their finger back and forth along the touchpad which other users within the VE can see if avatars are visible in the scene. The back button can be clicked to re-centre the POV of the user to the default position, and the home button will return users to a menu outside of the platform where settings can be adjusted. The system is meant to be used without headphones, allowing users to hear their instructor and communicate within the classroom normally. Sounds can be played from the connected phone or the instructor's device.

The system supports various types of content. These include 360˚ and regular videos, a tutorial room where 3D models can be displayed (both of which are controlled and manipulated by the teacher through their device), games in which students can collaboratively complete, PowerPoints, and panoramic images. The instructor can zoom, rotate and extrapolate models, pause/play and adjust videos, assign teams of users, track recorded data (i.e. game scores), and annotate onto any of the content types using a drawing tool (the line can be adjusted for thickness and colour). It is important to note that the above content was, at the time of this study, what was currently available to the platform. They should not be taken as representative of the full capabilities of the platform, as future content could utilise the characteristics of the hardware in different ways.

3.2 Theoretical Basis

As stated, CLT provided the basis for the evaluation criteria. By extension, guidelines from CTML further informed the creation of the criteria. As discussed in the literature review, it was concluded that the similar/superior effect on non-spatial cognitive learning outcomes of non-IVR and traditional methods, relative to IVR, are the result of less efficient schema development. This being due to IVR's tendency to produce large extraneous loads, assuming intrinsic load is sufficiently high (Sweller et al., 2011).

As stated in the review, this study focused on reducing extraneous load. CLT provides many recommendations for managing complex learning material and promoting deeper learning, but the sole focus for this study was the presentation of content.

To reiterate, a precise relationship between CLT and IVR has not been established. Nonetheless, it was concluded that unless the only intention is to provide a novel experience, it seems to be important to balance a low extraneous load with a sufficient level of immersion, even if the exact conditions are yet to be established. Firstly, implementing strategies that reduce extraneous load could reduce presence, but evidence suggested that the overall impact seems to be positive for cognitive learning outcomes. Secondly, the relationship between CLT and interaction & interactivity is unclear, especially for highly interactive content with 6Dof and peripheral devices of various modalities. However, interaction in the platform is relatively simply, involving selecting and moving objects using the touchpad and passive content, so the impact of its interactive nature on cognitive load is probably limited. The unique advantage of platform is its ability to provide immersive 360˚content simultaneously to many users, rather than complex interaction & interactivity. For these reasons an assumption was made that the primary difference between the immersive content provided by the platform and non-IVR multimedia formats is the 3DoF and the cognitive costs inherent to the limited FOV and navigation requirements of the HMD. Therefore, the principles CTML, which apply to dynamic media with low levels of interaction (Mayer, 2009), were assumed to be largely accurate for the platform's content, with some further affects due to the unique implications HMDs. The collaborative potential of platform also needed to be accounted for given its effects on cognitive load and the collaborative capabilities of the platform. Although collaborative principles of CLT have been established, as discussed in the literature review, they were not included in the synthesis of the criteria but were included in the discussions when relevant.

3.3 Research Framework

3.3.1 Method

As the research involved analysing and discussing visual & textual data, without quantitative results, a qualitative approach was deemed appropriate. The qualitative method of this study provided the data required to descriptively evaluate the software.

Qualitative research is the application of scientific methods that collect and analyse non-numerical data. More generally Miller describes it as an: "investigative process where the researcher gradually makes sense of a social phenomenon by contrasting, comparing, replicating, cataloguing and classifying the object of study" (Creswell, 2018). According to Creswell, qualitative research is defined by the core characteristics of: Natural Setting, Researcher as key instrument, Multiple sources of data, Inductive and deductive data analysis, Participants' meanings, Emergent design, Reflexivity and providing a Holistic account.

As stated, document analysis was the qualitative method adopted for this study. According to Bowen, document analysis "is a systematic procedure for reviewing or evaluating documents—both printed and electronic (computer-based and Internet-transmitted) material" (2009). The purpose of document analysis is to 'elicit meaning, gain understanding, and develop empirical knowledge'. 'Documents' can take a wide variety of forms (Bowen, 2009), but include visual sources and sounds (Denscombe, 2010) and instructional materials that accompany products (Neuman, 2014), which in principle includes IVR software like the subject platform. According to Bowen, document analysis is a combination of content analysis (organising information into research relevant categories) and thematic analysis (searching for patterns and themes within the data). For this study, document analysis was employed to identify where platform met the criteria, by a combination of analysing audio-visual material (directly examining the content) and textual documentation. Through this approach, the details of the content presentation were identified by defining the criteria first and using them to examine the IVR experience & documentation. In this sense, the data categories were defined before the data was collected, and the data was filtered through them.

The outcome of the research was a descriptive analysis of selected scenes with respect to the criteria, and descriptions of recommended changes/additions based off the criteria with respect to the context of the scenes. From this, a broader evaluation of the platform was to be completed based of the research questions (Q1, Q2), that is, how the characteristics of the platform help to reduce extraneous load defined by the criteria. Within each description was a summary of the scene, its intended learning outcome, where it met the criteria and where it did not meet the criteria, and possible changes. These are contained within the results and discussion (4). The evaluation is at the end of the results and discussion section. The dissertation was to end with a conclusion of the results and recommendations for future research (5).

3.3.2 Researcher's Role

As the researcher is the key instrument to the qualitative research process, it is important to identify the researcher's values, assumptions and biases (Creswell, 2018). At the time of this study I was completing a 'Masters of Teaching' and was a pre-service teacher. My approach was primarily theoretical & STEM focused, which could have introduced biases. Although every effort was made to mitigate the influence of these biases, they may still have shaped my understanding of and perspective on the data. As I was neither an experienced teacher or software developer, my perspective was likely less pragmatic/practical and more theoretically idealistic, being focused on the general recommendations of CLT. However, the evaluation was designed to be highly descriptive which could create room for the practicalities required for developers and teachers, as per the core characteristics of qualitative research. Of course, as a classroom educational platform, the intended audience of platform are school students, a demographic I am not a part of. However, the principles of CLT apply to school student and adult alike with prior knowledge being the primary delineation.

3.3.3 Data Collection

Given the evaluative purpose of this study, the platform itself, its content and supporting documentation was the source of all collected data. The platform is designed to display a wide range of content, with the intention of allowing teachers to create their own via the Creators Toolkit. In order to conduct a sufficiently broad analysis of the software, one scene from each of the following categories was examined:

- Captured 360˚ Video
- Animated 360˚ Video
- 3D model/Tutorial Room
- Gamified Scene

The specific scenes selected to represent the categories needed to be complete, the information they convey at a minimum understandable (video is of average resolution, dialogue is discernible, game instructions are clear etc.), and their intended learning outcomes at least partly cognitive.

The advantage of this approach is that any analysis applied to content within each category could carry over to other scenes within the same category. These categories cover the scope of the capabilities of the platform with respect to its content, with the intention of providing the necessary evaluative breadth whilst avoiding the problem of analysing an arbitrary selection of scenes or unnecessarily analysing them all. Furthermore, supporting documentation that describes both the platform and information specific to these scenes was analysed. This was done to ensure that the full range of the content was accounted for in case anything was missed during the examination of the scenes themselves. The scenes were treated as audio-visual material when examined, and the supporting documentation was analysed as ordinary textual document.

These data types have their limitations. Audio-visual material can be difficult to interpret, and the precise nature of the experience can be subject to the examiner's biases and deficiencies in their observational skills. The pre-defined criteria, which were designed to frame the examination, should have reduced possible bias, examining the scenes more than once would help to resolve observation related issues. Text documents can be incomplete and or inaccurate (Creswell, 2018). This was not considered an important issue as the documents were considered secondary to and supportive of the examinations.

Data collection during examination was done by taking observational notes and depended on the scene category:

- Captured 360˚ Video – Video watched till completion
- Animated 360˚ Video – Video watched till completion
- 3D model/Tutorial Room – All features of model and surrounding VE explored
- Gamified Scene – Game played until completion

These procedures ensured that the full range of content for each scene was examined. The observational notes firstly recorded where a scene in question met the criteria points, detailing which criteria was met and how. Secondly, the notes recorded where a scene did not meet the criteria, and how. The notes were recorded as events within the scenes happened and after the instructor's tools were tested.

3.3.4 Data Analysis

As discussed, the data was analysed against the criteria. In this way the criteria acted as a set of predetermined codes or 'qualitative codebook' (Creswell, 2018). The criteria were based off the effects/principles for reducing extraneous load in content presentation from CLT and CTML (Table 1 & Table 2), and are organised in Table 4:

Table 4: Criteria for data analysis

Code Label/Criterion	**Principle**	**Examples of Relevant Content Design**
C	Coherence	- Level of visual fidelity - Background sounds & music - Unnecessary text or narration
S1	Signaling	- Application of a cues - Object that provides visual guidance - Visual distinction of task relevant object
R	Redundancy	- Whether written and spoken words are non-identical and or not synchronised
SC	Spatial Contiguity	- Corresponding visual elements are spatially integrated - Pop up text - Capability of instructor/users to move related content closer and visually manipulate content
TC	Temporal Contiguity	- Corresponding audial and visual content presented simultaneously
S2	Segmenting	- Capability of users/instructor to pause and replay content - Pacing of content - Whether content is intrinsically segmented or continuous
M1	Modality	- Whether narration is used instead of printed text

M2	Multimedia	- Whether narration is supported by task relevant visual material
I	Image	- Whether the speaker of a narration is present on the screen

These criteria were applied when necessary, regardless of the category of the scene. However, from the literature it was expected that for each scene category, some criteria were more relevant than others. For captured 360˚ video, signaling, segmenting and coherence are especially important (Ibrahim, 2012). For animated 360˚ videos, segmenting, modality, contiguity and signaling (Khalil, Paas, Johnson, & Payer, 2005; Wouters et al., 2008). For the tutorial room and 3D models signaling, spatial contiguity and modality are likely to be significant, given the largely passive nature of the category. Recommendations from CLT for educational games are less clear, but it was expected that collaboration would be most relevant for this category.

It was important to consider the boundary conditions for these criteria, that is, the conditions under which they would be most relevant. The important conditions are 'individual differences', methods that work for novice learners may not work as well for experts (which is the same as the level of expertise in CLT), and 'complexity and pacing', which states that the methods of CTML will be more relevant for complex and or fast-paced content (Sweller and colleagues describe the same effect in the language of high intrinsic load) (Sorden, 2012). The former condition depends on teachers' choices, as the target audience of the platform is school students (above the ages of 12/13) there is not going to be much variation between their levels of expertise. If there is, it is the teacher's responsibility to create and or choose content to show their students accordingly. Therefore, this condition was not considered to be important for the study as the scenes were evaluated on the assumption that the students viewing it would have the appropriate prior knowledge.

The latter condition on the other hand was incorporated in the evaluation. If a scene, or part of it, was evaluated to be low in intrinsic load then most of the criteria are not necessary. As identified in the literature review, the design choices demanded by the criteria could reduce the authenticity of a scene and in turn reduce its immersion and capability to establish presence. As immersion and presence have been associated with strong affective learning advantages, introducing design choices that unnecessarily reduce extraneous load and potentially reduce immersion and presence will do more harm than good. Estimating intrinsic load can be difficult given that it is dependent on a student's prior knowledge and element interactivity. Given that for school students even tasks with seemingly low intrinsic load can result in comparatively lower learning outcomes in IVR (Feng, 2018;

Parong & Mayer, 2018), only slow paced content with single elements presented at a time was considered to be low in intrinsic load.

This connects to the intended learning outcomes of a scene, all or part of it. If the intended learning outcomes are affective then it is more important to increase immersion and presence than reduce extraneous load. However, the criteria will not affect immersion and presence under the right conditions, making them ideal for reducing extraneous load when affective processing is high. In general, if the implementation of a criteria recommendation is done in a way that does not affect the authenticity of a scene, then immersion is unlikely to be affected. Furthermore, in general, if the cognitive load of a scene is so high that there is insufficient space for affective processing, reducing extraneous load is necessary even if the intended learning outcome is affective.

A summary of the analysis process is described by Figure 4.

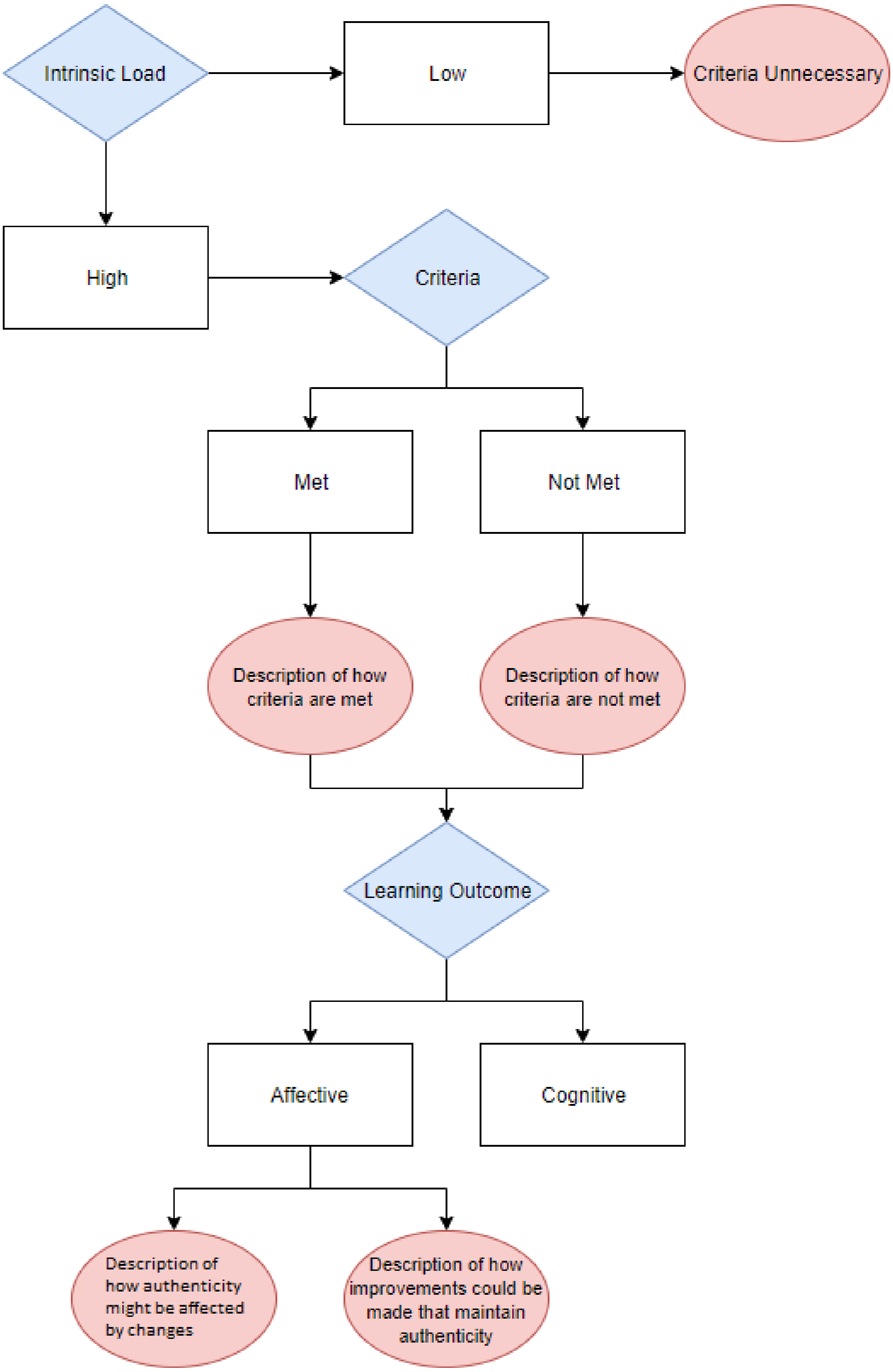

Figure 4: Overview of analysis

3.3.5 Validity & Reliability

Several methods were used to improve the qualitative validity & reliability of the study, following the recommendations of Creswell (2018). For validity, the results were conveyed using extensive descriptions which allowed for details, exceptions and nuances to be incorporated as part of the analysis. The researcher's background and potential biases have been noted, giving context to the interpretations made. For reliability, the codes were to be cross-checked by another researcher who independently applied the criteria to the examination notes and documents.

3.3.6 Ethical Considerations

The study worked exclusively with the platform's company, no students were involved. A Memorandum of Agreement was signed such that informed consent, voluntary participation, agreed upon level of confidentiality, and a right to withdraw were ensured. Furthermore, the memorandum included an agreement that the results would be accurately, fairly and constructively reported on. At no point were the results published or reported without the consent of company and the research supervisors.

3.4 Limitations

There were several notable limitations in the study. As stated numerous times, a precise relationship between the implications of CLT and the unique affordances of IVR has not been established. More generally, the relationship between germane processing and motivating pedagogies is under development (Mayer, 2014). Given the unique capability of IVR to promote affective outcomes, the balance between improving affective outcomes and minimising cognitive load was relevant to the study but was an open question. The novelty of IVR, or the amount of prior IVR experience students have, can influence learning outcomes as well (Makransky et al., 2019; Markowitz et al., 2018; Moesgaard, Fiss, Warming, Klubien, & Schoenau-Fog, 2015). Novelty was not factored into the study. The study focused on reducing extraneous load, design recommendations related to managing intrinsic load and promoting germane processing were not considered. The study relied on a subjective approach to evaluating the platform. Several measures had to be made to minimise the influence of this fact. Finally, the platform had a broad potential and versatility, this could only be covered by analysing a few scenes, which could be limited in coverage.

4 Results & Discussion

4.1 Captured 360˚ Video

4.1.1 Video Selection & Description

A range of captured 360˚ videos were available to analyse. The 'Barefoot Forest' series was selected. Barefoot Forests consisted of 4 stereoscopic videos (also available without stereo), each over 2 minutes in length (for a total of 6 minutes), in which an instructor guides viewers through a forest in NSW. The videos are broken into segments which are either narrated or with the instructor in view. During these videos the instructor briefly explains a range of key ecological concepts including different trophic levels in the food chain, photosynthesis, adaptation, and ultimately how these concepts contribute to the function, form and development of a forest. The stereo quality adds visual depth to the videos but with a reduced resolution relative to the non-stereo alternative.

This video series was selected for several reasons. The videos available were designed for a range of demographics, this series in particular was designed for (primarily) school students for describing basic & fundamental ecological concepts. As the series covered a reasonable length and a range of concepts, a larger variety of content was available to analyse and compare. The videos also adopted a few presentation styles with some segments narrated and in others the instructor acting in the shot. Lastly, the videos were complete, including intro and outro segments, and of a high quality without technical issues.

4.1.2 Intended Learning Outcomes

The intended learning outcomes of the videos were a mixture of cognitive and affective. The video was meant to teach users several ecological concepts, but was also intended to simulate the experience of walking through a forest, hence the title of the series. It was important that users feel like they are walking through a real forest, so the design methods used to reduce extraneous load should avoid impinging on the authenticity/realism of the film. Changes that are not natural to the setting will harm this feeling of authenticity, such as graphical overlay effects or artificial sounds.

4.1.3 Criteria: Analysis

4.1.3.1 C

The video was of a high fidelity and quality, assuming that the HMD has been properly adjusted for the viewer. Small details could be made out several meters from the camera, beyond this distance theses details became less visible. The effect of video fidelity/quality on cognitive load is not clear although it is possible that, if it is low, users could spend resources trying to make out blurry images. In any case higher quality videos will clearly elicit greater immersion, therefore high fidelity is recommended for captured video. With one exception, important visual elements were easy to make out in the video. The sole exception was that at one point the instructor picked up a mushroom to show, however the mushroom is difficult to see as the instructor is standing too far away which could be easily alleviated by moving sufficiently close to the camera.

Natural, background sounds of the forest can be heard. As these sounds add to the authenticity of the video, they could be considered necessary for the intended learning outcomes. Furthermore, they are never distracting or louder than the instructor. In general, it is recommended to remove sounds that are not explicitly relevant to the learning outcomes. If the spoken text of the video was lengthier or more complex, or the intended learning outcomes were primarily cognitive, either removing the backgrounds sounds or reducing their volume when the instructor is speaking would be necessary.

4.1.3.2 S1

For most of the video, visual and auditory cues were unnecessary since pointing out specific visual elements was only intended in a couple of segments. Signals that were identified included gesturing and auditory orienting-cues (or directions). The instructor used a range of gestures to guide the user to relevant visual material. In one segment the instructor refers to several species of plants by moving towards and interacting with them. Otherwise the gestures are simple and brief which was appropriate since the instructor was speaking about general visual features.

Likewise, general auditory directions were spoken, such as asking the user to look around or describing/directing-towards important visual elements, more specific cues were not necessary at any point. If the video intended to discuss more specific visual elements then gestures and spoken orienting would be ideal as, unlike artificial cues (arrows, flashing, colour changes etc.) they won’t populate the video with unnatural elements that might harm its authenticity. However, just as general cues work for general features, specific features require specific and clear cues, for instance ensuring that a gesture guides towards a clear direction or precise location, or that the orienting cue describes the location of an element relative to another one.

As spoken text ran throughout the video, stressing key words is crucial for distinguishing them, which the instructor did. This cue shouldn’t influence authenticity and is generally recommended in any situation in which spoken text is used.

The drawing tool could be an effective signaling tool. The instructor could draw arrows, circle critical elements, number elements etc. The advantage of this tool is that a instructor is free to create and place signals like these, if they do not want to affect the immersion of the video they can simple not use it. One consideration of this tool is that it seemed to pause the video when used which might not be preferable at the moment when the instructor would like to use it. Having the option to use it without pausing the video could make it a more effective tool for signaling.

4.1.3.3 SC

Given that specific visual elements were not often focused on and no written text was used, there was a limited need for spatial contiguity. A good design choice used throughout the videos was that the important information was always contained within the default FOV, at no point in the videos was the user asked or required to look for specific things at extreme angles. This included the location of the instructor who stayed within the initial FOV and, when necessary, stood next to important elements. The drawing tool could be helpful for improving spatial contiguity. For instance, it could be used to visually connect or label critical elements.

4.1.3.4 S2

The videos are segmented into relatively brief segments. Asides from changing the shot this segmentation makes it easier to determine when to pause to ask users questions or ensure that they absorbed what was said. The transitions are brief fade-to-blacks giving time for the instructor to pause if needed. The videos are not user paced as the content is controlled through the instructor's device (which is a good thing considering that the users are watching the same video). It is up to the instructor to determine when to pause or replay content and communicate with users. The narration is paced slowly, but the intrinsic load can be high at times since many different concepts and their terms are discussed, although not in detail. For less knowledgeable users pausing and replaying (after communication) would be recommended.

4.1.3.5 M2

For quite a few shots the visual content and the narration do not match. It is possible that users will be distracted from listening to the content discussed by looking around, however this can be difficult to avoid given the learning intentions of the video. Discussed concepts could be matched with close ups of examples in the forest or simply editing images in, however for these videos this could take away from the feeling of walking through a forest. If this extraneous load hard to remove, reducing intrinsic load will be required.

4.1.3.6 R, TC, M1, I

There is less to say with regards to the remaining criteria. R/M1: No written text was used which is generally recommended, furthermore the videos never focused on specific audial elements which otherwise would best be paired with written text. TC: Specific corresponding visual and spoken information was temporally contiguous, but otherwise irrelevant as specific visual elements were rarely focused on. I: When on screen the instructor was meant to be the centre of attention to guide the viewers and add to the authenticity, in these moments the instructor's presence was necessary.

4.1.4 Discussion

It is clear that the design of a captured 360˚ video is highly dependent on the intended learning outcomes, a result that was found by Ibrahim (2012). For the Barefoot Forests videos, it was critical that authenticity of the video is maintained, users are supposed to feel like they are taking a tour through a forest. Hence, any adjustments to reduce the extraneous load of the video should not reduce the realism of the scene. Overall, the videos met the criteria well and elicited a strong sense of presence, but several possible adjustments were discussed. By analysing the Barefoot Forests videos, several important considerations were found for captured 360˚ videos in general.

The quality of the video needs to be considered when filming it, if a specific visual element is discussed it needs to be clear enough to identify. It is possible that users will be distracted from a task by trying to make out a blurry object, creating a source of extraneous load. A limitation of captured video is that there will always be visual and audial information that is interesting but irrelevant to the specific learning tasks, which will create a redundancy effect. This issue will increase with the complexity and the filmed environment and as the visual elements become less distinct (Olk et al., 2018). Removing redundant elements with editing could significantly affect the realism of the video and possibly its motivating effects (Ibrahim, 2012), rather a thorough use of alternative methods to reduce extraneous load is needed, primarily diverting attention to the relevant elements and concepts.

Spoken text is always recommended over written text due to the modality effect and a removal of split attention, unless an audial element (a specific sound or set of sounds) is central to learning. In this case simply switching between the two when appropriate, but avoiding using both simultaneously, will do. Otherwise the 360˚ environment is so visually demanding that using written text will be distracting. This also means that spoken signals will always be useful, stressing key terms and descriptively guiding the attention of users with auditory cues should be staples for captured 360˚ content. This is not limited to recorded sounds, without headphones the instructor can speak to users normally so, if needed, the instructor can use spoken signals of their own. The effectiveness of these signals will depend on the instructor's clearness, timing and communication with the users. When the required search is low, it is generally recommended to uses an auditory cue over a visual one (W. Roberts, 2008).

As discussed by Feng (2018), navigating a 360˚ environment can be time consuming and requires cognitive resources which can potentially frustrate and disengage users. Likewise, it is important to keep users focused in the correct parts of the video and reduce wandering. Visual cueing is a versatile method to reduce search time but depends on the intended learning outcomes, including whether specific visual elements or locations are going to be referred to. Clear cues can be edited onto the captured video but if this isn't preferable an instructor using gestures could be an effective alternative. It is important to note that the impact of gesturing on learning is still unclear (Davis, 2018), but without editing there aren't many alternatives and, given that 360˚ videos can have particularly large distances between elements having some sort of signal is preferable to none. Visually distinguishing important elements can also help reduce search time. However, this is dependent on the environment filmed as well as the intended learning outcomes. Sometimes specific elements will be already visually distinct (like a colourful plant), otherwise the environment will need physical adjustments or editing such that the important elements are made to be distinct. Finally, ensuring that the important or mutual dependent visual elements are close to each other (within the FOV) will also reduce search times. As discussed with Barefoot Forests videos, containing most of the important elements within the default FOV is a good standard. Users have a wide space to navigate that, at large angles, will require them to rotate their heads (with devices) to a great extent, beyond which they would have to rotate their bodies entirely which could be uncomfortable, time consuming and distracting.

Segmenting videos is the last important consideration for captured 360˚ video, and for the same reasons as regular video. Segmenting helps compartmentalise a video's information into more manageable parts and will reduce the need for users to identify the topical boundaries themselves (Ibrahim, 2012). Following this the transitions between segments are also a good time for the instructor to pause the needed. Segmenting allows for a change in scene such that the visuals are more relevant to the learning task, or to change the direction of the camera to a more relevant view of the scene rather than asking the users to rotate. As the Modality effect disappears when spoken text becomes to lengthy or complex (Sweller et al., 2011), segmentation will help to keep it within bounds.

These considerations come down to the production of the videos and how they used in the classroom, the platform itself is well suited to accommodating this type of scene. A typical problem for HMDs is that they are often meant for standalone use (Jensen & Konradsen, 2018), normally it would be difficult for an instructor to manage a large number of users undertaking their own self-contained HMD experience. It could be difficult for a instructor to check their user's understanding or guide them if they are having trouble navigating the video. The capability of the platform to support many users, and the tools it provides a instructor to control the and interact with videos, puts it at a distinct advantage. As each user watches the same video stream, the instructor can keep track of the entire class, which means that signals (such as drawings or classroom communications) and segmentation (pausing and replaying a video) will apply to the whole class simultaneously. Despite having control over a video, the instructor can still involve the users by communicating with them normally i.e. asking them if a pause or direction is needed. The drawing tool is particularly useful as it allows the instructor to create signals and spatially integrate elements when the situation demands, the fact that the instructor can easily communicate with the class makes it easier to determine when a drawing is needed. It should be noted that without headphones it is possible that a user could get distracted by sounds from the classroom, creating extraneous load and reducing their sense of presence. It is up to instructors to appropriately apply the platform and communicate with the class in such a way to reduce these possibilities.

4.2 Animated 360˚ Video

4.2.1 Video Selection & Description

The ‘Biology Tour’ video was selected among the few animated videos available. The video takes place with a 3D simulation of the inside of an animal cell and is 6 minutes long. The complex video includes many cellular components and their processes occurring simultaneously, including organelles, proteins, mRNA translation, transportation of vesicle, polymerisation of microtubules, and so on. The models are highly detailed with realistic animations and are accurately scaled, yet they are also colour coded and not overly detailed so that the environment is easier to understand. As the video runs these components and processes are described by a narrator, including their wider effect on the cell.

As with the Barefoot Forest series, this video was selected because of its length, quality, partly cognitive learning outcomes and because it was complete. The video covers a lot of concepts though narration, so the intrinsic load is relatively high. Furthermore, school students are the intended demographic. The video is also a good example of size, transduction and reification allowing the analysis to include these notable features.

4.2.2 Intended Learning Outcomes

The intended learning outcomes of the video were both cognitive and affective, but this could depend on how the video is used in class. The video described and demonstrated many cellular concepts so there were clear cognitive learning outcomes. On the other hand, the video could be used to prompt interest in the topic or provide a concrete representation of concepts previously described in class, used in this way the cognitive outcomes are less important. As the video was an animated representation of a cell the effect that changes might have on authenticity were assumed to be less important. Some of the design choices of the video were already designed for learning over realism, such as the colour coding and differing levels of detail for the cellular components.

4.2.3 Criteria: Analysis

4.2.3.1 C

The visual fidelity of the scene was very high as the models were easily distinguishable and autonomous, with undulating surfaces and realistic movement. The fidelity was much higher necessary to distinguish different objects and understand the processes, from a cognitive perspective much simpler models would be enough to build the conceptual understanding. The extra detail comes down to the intended affective outcomes of the video, as the VE would certainly not be as immersive if the animation was more simplistic and less realistic.

The video included many moving objects and their processes, all occurring mostly simultaneously. When the narrator was discussing one specific process many others that were irrelevant to dialogue were also visible. Being irrelevant, these processes could be distracting and contribute to extraneous load. Again, these processes occurring simultaneously is more realistic and would likely contribute to superior affective outcomes, including or removing this quality depends on the intended learning outcomes of the video. Similarly, the video includes a constant ambient sound, although it is atmospheric it is also irrelevant to the concepts explained and could be a minor distraction.

4.2.3.2 S1

All of the identified cues were auditory cues from the narrator. The narrator guided the user’s attention by describing the shape and colour of certain objects and giving basic directions to look in. Most of these cues were basic and didn’t take too long to say. The issue is that the environment was so complex that sometimes these cues needed to be clearer so that users could identify exactly what object or type of object the narrator is describing, especially when the narrator asks users to look in a certain direction first. The narrator did not stress key words, which is a change that could be easily made. The only visual signaling evident was the use of colour coding, this was very useful given absence of visual cues as it gave the narrator more information for descriptions. As an animated video, there are far more opportunities for visual cueing as the cues can be built into the animation, for instance objects could flash or change colour, the background could change, guiding objects like arrows could be modelled in etc. As with captured videos, the use of visual cues depends on the learning outcomes of the video and the level of realism established. Some, like arrows, will probably reduce the realism of the animation as they are unnatural to the environment, others will have far less of an effect such as colour changing, as most objects in the animation are already unrealistically coloured.

4.2.3.3 SC

As the processes described were animated, they were naturally spatially integrated. An effective design choice used was that as the video changed concept, it tended to discuss an object or process adjacent the one previously discussed which will help reduce search times and reduces the need to cue. There were no points in which mutually dependents visual elements were significantly separated in space, as the narrator only discussed objects that the users would presumably be looking at. The drawing tool can be used to draw connections between visual elements if needed, as with the captured videos.

4.2.3.4 S2

The animations and narration ran at a medium pace. The video needed more breaks between exposition, especially when users are asked to look for certain objects. The instructor still has control over pausing and replaying, but making the breaks slightly longer would give the instructor more time to pause if needed particularly if they want to communicate with the users first. The video reset the scene on a couple of occasions to refer to discuss different processes that were visible earlier, but not focused on. This included a brief fade to black and provided clearer boundaries for segmentation. As the narration was concept heavy, pausing and replaying would be strongly recommended for less knowledgeable users.

4.2.3.5 R, TC, M1, M2, I

R/M1: No written text was used, neither was written text preferable as the only sounds in the video was the narrator and a soft background sound. TC: Mutually dependent narration and visual content were simultaneously presented. M2: Most of the narration was supported by visual content, on a couple of occasions the narrator discussed processes that were not visible (such as the processing of proteins in the golgi body). However in these cases, unless the concepts are going to be discussed in detail, the multimedia effect (M2) is not important as it describes exploiting more WM capacity which is needless if the intrinsic load is within the limits of one channel. I: The narrator was not visible which is recommended.

4.2.4 Discussion

The video was good example of the unique capabilities of IVR, as there is simply no other way to experience the inside of a cell with such immersion, however several places for potential improvements were identified. The video would benefit from longer breaks between exposition and when users are given directions, and the use of visual cues and stressing key words would further help users identify critical information. The level of visual fidelity and the density of objects and processes depends on the learning intentions of the video, to improve cognitive outcomes both could reduced. Some general characteristics of animated 360˚ video were noted after analysing the Biology Tour video.

Many of the principles that apply to captured 360˚ videos apply to their animated counterparts. Animated 360˚ videos benefit from spoken text, segmentation, a range of cues and integration. The most significant difference is that the animators have significant control over the environment and subsequently many more options for reducing & managing sources of extraneous load. This is especially true for coherence, where animators can precisely control the level of visual fidelity whilst also not having to worry about the technical issues that come with captured film. The quantity of objects within a 3D environment is easier to control as objects within the environment can added or removed at will during a video's creation, whereas for captured videos the environment must be physically manipulated, or difficult editing techniques would have to be used. For minimising cognitive load and maximising cognitive outcomes, animators should aim to minimise the visual detail of an animation whilst maintaining the required distinguishability of objects & their processes that necessary for the learning task. As a simple example, in the Biology Tour video the undulation of the organelles & proteins would only be a relevant detail if understanding the fluidity of these objects (or something similar) is an intended learning outcome. Otherwise it might contribute to immersion but could distract from the learning task.

If possible, it is best to rely more on visual information than text, the advantage of animations is that they serve as dynamic visualisations. Watching a process removes the need for mental inference and frees up WM resources, and many processes can be difficult to verbalise (Ibrahim, 2012). It is also important to avoid explaining a process that can clearly be seen due to the redundancy effect, instead using dialogue for naming and dropping key terms for the visible processes. Temporal incontiguity is acceptable if the text is segmented and the offset is only a few seconds long (Wouters et al., 2008). As an educational tool, animations are suggested to be superior to static visuals under the apprehension principle (essentially segmentation/pacing) and the congruence principle (Morrison, Tversky, & Betrancourt, 2000). According to apprehension principle, animations must be easily perceivable and properly paced. Each part of an animation will take time to process in WM to be integrated into LTM. If an earlier part of an animation is not stored before subsequent parts begin processing, a user will struggle to build a coherent representation of the concept animated since each part is mutually dependent and difficult to understand without the others (Wouters et al., 2008). The congruence principle states that the structure of an animated process should correspond to how people tend to conceive the process.

Animated videos are better suited to visual cues for various reasons (W. Roberts, 2008; Wouters et al., 2008), but it is also worth noting that they are less artificial in an environment that is already fully animated. With captured video there are few ways to add visual cues without editing them in after filming, an on-screen instructor being one of the few examples. Colour coding or colour cueing is one of the simpler and possibly less intrusive visual cues. Colour can be manipulated in several dimensions which is worth taking advantage of, should be limited to a maximum of six different colours, and the more important the object the brighter its colour (W. Roberts, 2008). However, evidence suggests that the significant effects of colour coding are lost when the colours do not correlate with the information they denote, that the brain processes different colours differently, and colours tend to be integrated into mental models causing extraneous processing when the colour scheme is changed. Animated instructors, or pedagogical agents, are an option as well, but the research is mixed so they should be applied with care (Wouters et al., 2008).

The main downside to animated videos is that under most conditions they will be more expensive to create than captured videos which creates a greater impetus to ensure that from an educational standpoint they are designed optimally.

As with the captured videos, these points are entirely production based, the platform is ideal for managing the content. This is for the same reasons discussed for captured video, the platform, when using GearVR, only limits interaction which is mostly irrelevant for video of any type.

4.3 3D model/Tutorial Room

4.3.1 Video Selection & Description

The ‘Tutorial Room’ scenes are 3D VEs which includes the users’ avatars and usernames. The content of the scene is still controlled by the instructor, asides from waving to each other users cannot directly interact with it so the user experience is still passive. To reiterate a point discussed in the methodology, unlike with the videos the exact characteristics and limitations of these VEs is difficult to quantify. The platform could support many possible designs so long as it shares the primary characteristics of being a 3D VEs which includes avatars, limited user interaction and paced by the instructor. Of course, to conduct the analysis the rooms available were selected from, which shared more specific design characteristics in common. In these scenes the instructor can display various models which include inbuilt options such as paling a fixed animation, blow out the model to display parts, displaying labels etc. The instructor can rotate and zoom these models and use the drawing tool to further interact with the scene. The rooms varied in complexity, most were a circular platform with avatars distributed in a circle, but more complex rooms were available such as an anatomical theatre and an office room.

From the rooms available the ‘Cell Tutorial Room’ scene was selected. The room is a circular platform with three different eukaryotic cells models floating in the middle. These models are cells cut through the middle, with the important organelles and other components visible and colour coded. The instructor can select one of the individual cells or a side-by-side of two cells for comparison, they can also bring up labels which name each component of the model, with the component and label connected by a line drawn between them. This scene was selected because it featured a range of models with options, was intended for school students and had clear cognitive intended learning outcomes.

4.3.2 Intended Learning Outcomes

The scene was designed with cognitive learning outcomes in mind. Asides from the intrinsic affective qualities of using an HMD, the scene was clearly designed to help users understand and visualise the different parts of cells, and for understanding the differences between types of eukaryotic cell.

4.3.3 Criteria: Analysis

4.3.3.1 C

The level of visual fidelity of the models was appropriate given the learning task. The models were completely static, with some textures which helped depict the nature of the textured surfaces. Given that the learning intentions of the models was to show the types of structures within cells and where they are positioned, further detail beyond the basic shape of these parts would be unnecessary. Only one or two models would be clearly visible at one time, which helped to reduce the chance that user would focus to task irrelevant material. The background of the scene included drifting cell-like objects which could be unnecessary for learning. Given that the scene is paced by the instructor it is unlikely that the extraneous load caused by these objects would be detrimental, as users would have plenty of time to divert their attention toward the models if distracted and the intrinsic load of the scene is not high. They also contribute to the atmosphere which could improve the immersion of the scene. The scene did not include any background sounds, music or dialogue which is suggested.

4.3.3.2 S1

A couple of signals were observed. The cells and their components were colour coded which helped to distinguish the parts in what were relatively crowded models. As the models were static and scene instructor-paced the drawing tool is particularly useful for pointing out critical information, as there is no risk of disrupting a dynamic process.

4.3.3.3 SC

The scene made good use of spatial contiguity. When labels were displayed the labels and the components they refer too were connected with solid lines. As the models were quite densely packed, this approach is better than fully spatially integrating the labels against their respective components which would overly clutter the model. When a model was selected by the instructor, all the users are transported to a shared POV facing the model. This immediately made the model the centre of attention and removed the need to search, and would also ensure that each user in the class is looking at the same thing (assuming they are looking forward). The models could be zoomed in & out and rotated by the instructor which could help move parts of the models closer to the POV of the users. The labels are a part of the model and would maintain their position relative to the model when these manipulations were made, this did mean that these labels become unreadable if the model is rotated beyond a certain angle. This could be avoided if the labels dynamically adjust to face the users as the model is rotated.

4.3.3.4 R, TC, S2, M1, M2, I

R/M1/M2/TC: Only written text was used (in the form of labels) so these criteria were irrelevant, verbal information is entirely dependent on the instructor. S2: The scene had no intrinsic pacing and was entirely controlled by the instructor. I: There was no speaker/narrator, and the instructor does not appear in the scene which is recommended.

4.3.4 Discussion

The scene followed the criteria well with a few small areas of improvement, however many of the criteria were not sufficiently relevant. The scene added more user involvement than the videos and did not have their intrinsic pace, which is the most likely reason these criteria were less applicable. It is still possible that a scene, that could still be considered a tutorial room, could include content in which these criteria would be more relevant. For instance, a tutorial room could have short narrated animations, the instructor could have an avatar present in the scene, etc. Hence, it is difficult to generalise the analysis of this scene to other scenes of the type (that the platform is capable of supporting). However the tutorial rooms available did have many important features in common which they shared with the Cell Tutorial Room.

One significant commonality was that the intrinsic load of these scenes was not very high, primarily because of the control the instructor had over the scene's pacing, and the low element interactivity of content that they were designed to teach. This means that the redundancy effect and the modality effect will be eliminated if the instructor correctly segments, giving users enough time to search for and integrate information (Sweller et al., 2011). Instead, providing the instructor more methods to reduce split attention and direct the attention of users is important. When used in a classroom, these scenes will also be centred around the instructor's dialogue which will instruct and guide users during the scene, this influences the suggested design further.

As the scenes are primarily meant to produce cognitive learning outcomes, minimising their visual complexity important. Keeping the models static is recommended unless that animation depicts a concept that is relevant to the learning task, with the exception of an optional animation or similar feature. Due to the instructor's need to speak during the scene, removing unnecessary sounds will not only remove a source of extraneous load but will remove a sound that might compete with their voice, this includes spoken text.

Written text, especially if toggleable, will perform better in these scenes. This is because these scenes are instructor paced and, assuming the instructor provides enough time, users can re-read complex text if they need to and skim over the simpler parts (Sweller et al., 2011). Pop-up text could be useful as not only will it add more user interaction, the text will appear automatically spatially integrated with its referent and will disappear when a user turns away. This could be implemented quite easily by appearing when a user's cursor passes over an object, or perhaps after the user looks at the object and taps the touchpad.

A significant consideration of these scenes is that the instructor is invisible to the users (both in the scene and in general due to the coverage of the HMD) but central to the content. This means that signaling is the most important principle for these scenes as the instructor needs ways to divert the attention of the users towards the elements they intend to teach. Visually, the drawing tool is extremely useful for this purpose, but there are many other possible ways to create visual signals that are not implemented. The instructor could tap an object and cause it to ping or flash, place a text box, the users could have a navigation device built into their UI that tells them in what direction to look, etc. This is not to say that such implementations are necessary, but as a tutorial scene becomes more complex providing the instructor more options to divert attention could be helpful. Audial cues and directions are probably unnecessary since instructors could create them themselves. As discussed in the animated video section, colour coding is a simple and obvious way to distinguish parts of the models.

For the same reasons as the videos, the platform suits this format well. Most of these suggestions apply to the creation of these scenes, rather than changes to the platform itself. The platform's capability of supporting multiple users as well as an instructor, allows multiple users to observe the same content whilst an instructor manipulates that content for instructional purposes. The platform could benefit from more tools for the instructor to use to create visual signals to direct the attention of users. The drawing tool was an example that was already implemented and functioning, however it has its strengths and limitations. It is quite versatile in that any basic 2D image can be drawn, however it does not alert users since they will only see the drawings when they look in their direction. A signaling tool that affects users no matter where they are looking, and helps them navigate to a desire direction, could be helpful.

4.4 Gamified Scene

4.4.1 Video Selection & Description

The platform supports games thanks to the touchpad on the side of the GearVR HMD and the platform's capability of supporting multiple users in the same environment. Users can look at objects in the environment and interactive with them using the touchpad, this interaction could be anything from picking the object up, selecting it to answer a question, selecting it and selecting on another object to associate them in some way, and so on. Since multiple users can occupy the same VE, collaborative possibilities open up as users can select the same objects and conduct ordinary discussions between each other within the classroom. An instructor can set up teams simply by selecting usernames and assigning them to teams, or the system can divide users itself. The platform also records and stores data, for games this could be scores and time played and the game can include real time feedback for users. This data can be used to establish benchmarks that future players can be compared against. The games that were available involved selecting and placing objects, two of which supported collaborative teams. Each of the games differed slightly but avatars were visible in each.

The 'Cell Game' was selected as it was complete, designed for cognitive learning outcomes, was collaborative with different roles, and was intended for school students. Users can play in pairs to answer questions and place organelles and other structures into a 3D model of a cell. One user receives a question about a cellular structure which they use to tell their teammate. The teammate has a series of answers (which are the names of the structures) adjacent to them that they can select from, they can also pick up one of the structures displayed in a row in front them and place it within the cell, hopefully the one described by the question. Both users can see the cell and the objects that are placed in it, so they can work together to ensure that the right structure is placed in the cell before the answer is selected. The room looks like the cell tutorial room scene, with the users standing on a circular platform, facing the edge to see the cell model in front of them. A wall divides the players so that they can't see each other directly. Questions and answers are displayed on panels suspended on the flanks of the players.

4.4.2 Intended Learning Outcomes

The game was intended to help users remember different cellular components, their purpose and their location within a cell. As a game the scene was likely to intrinsically support affective outcomes, but asides from the gamified design elements the outcome was clearly cognitive.

4.4.3 Criteria: Analysis

4.4.3.1 C

The visual fidelity of the scene was the same as the Cell Tutorial Room, with the same room and models used. This was appropriate for the tasks since users needed to be able to distinguish between the different components and read the questions and answers, but nothing further. Like the Cell Tutorial Room, there was background sounds, music or dialogue, and the written text was only used for the questions and answers making it necessary for the task.

4.4.3.2 S1

Answers and components illuminated when selected, providing visual feedback. No audial signals were used, a brief audial signal as feedback could be useful to alert users to when an answer changed. The questions and answers were positioned on the users flanks which were difficult to notice at first unless the user had taken some time to look around. It could be helpful to guide the user's attention towards these areas when necessary, such as when after a component is placed in the cell. However the scene is entirely user-paced, so once users are familiarised with the VE signals could be less impactful.

4.4.3.3 R, SC, TC, S2, M1, M2, I

Most of other criteria were less relevant for the scene. R/TC/M1/M2: No spoken text was used. SC: The scene was entirely user paced so segmentation was unnecessary. I: No instructors or animated agents are present in the scene.

4.4.4 Discussion

Many of the criteria were not directly relevant to the scene, there could be many reasons for this. As discussed earlier, as with the Cell Tutorial Room, this scene was not entirely representative of the games that the platform could possibly support. The platform could support games with narration, pedagogical agents, with multiple sections etc. For instance, if a game scene did have narrated sections then the criteria associated with spoken text (R, TC, M1, M2) would have been relevant. Furthermore, the scene itself was entirely user paced, so segmentation was completely irrelevant. If there were timed components, tutorial sections or similar temporary sections where information was imparted, then segmentation would have been important. However, the game scenes in general were also the only type in which collaboration was important, which was not strictly a category and is worth discussing. Finally, CLT and educational video games, or gamification in general, is itself an area without much research. The reasons for this are very similar to CLT and IVR, games favour interaction and motivation (Alsawaier, 2018) but cognitive load is still relevant, with higher cognitive loads correlating with poorer game performance as shown in one study by Beserra et al. (2014).

Games are complex learning environments, with presentation, interaction, collaboration, difficulty and many other factors that learning success will be a combination of. Trying to determine precisely where the sources of cognitive load are can be difficult, especially if a game is played within immersive VE.

With respect to coherence and signaling, the two criteria for which observations were made, the suggestions made for the tutorial rooms more or less apply to the gamified scenes. Removing extraneous load, that is, minimising the visual and audial complexity of the scene whilst maintaining enough content for the intended learning outcomes, is critical if those outcomes will be primarily cognitive. This point is echoed by Beserra et al.: "…all elements present for the player on the display should be considered, because while processing all the information present, the interaction of the player with the game can become difficult" (2014, p. 352). Signaling will be necessary for guiding the attention of users in the 360˚ environment and providing feedback for their actions. Unlike with the tutorial rooms, the instructor will have less influence in these scenes so automatic signals are most relevant, rather than those controlled by the instructor.

Collaboration can differ between games. In the Cell Game teams consisted of two users who had a different role, this was not the same for all the games the platform supported or could support. In one game users had the same role in a larger team, in another users played as individuals. After analysing the cell game, and looking at some of the other games the platform had, there is no clear reason why the collaborative CLT guidelines provided by Kirschner et al. (2018) will not apply to the IVR environment. The complexity of the game will dictate whether or not teams are necessary and how large they should be, if the game is not complex enough extraneous load can arise from unnecessary communication between the team members. Making a game sufficiently challenging is also necessary for engagement (Alsawaier, 2018), but too complex and the players will lose interest. Similarly, as the skill and knowledge of a team's members becomes more heterogeneous their transactions will incur further extraneous load. The platform has capabilities to deal with this as it allows an instructor to create teams & assign members and track relevant data. Using the platform an instructor has control over team creation and receives feedback for this from tracked data, this helps an instructor creates teams with the right composition and suited for the challenge. Different roles, as in the cell game, are generally recommended as their responsibilities are clearer, this reduces the risk of extraneous load emerging from disputes and confusion regarding responsibilities during transactions. Task guidance can also reduce extraneous load, the platform is capable of this as the instructor can observe users from their device and provide guidance through dialogue or visual cues using the drawing tool. The instructor could observe different teams by switching between them from their own device, whether the platform had this capability was unfortunately not discovered however it could easily be added if it wasn't already. The size of a team depends on the task but in general extraneous load will increase as team sizes increases. For the Cell Game the task wasn't complex enough to justify a third team member, so two player teams was probably optimal.

An interesting corollary is that these collaborative games are unique in that that are essentially an IVR version of a 'Shared Display'. In a shared display game, multiple users share the visual display and play from their own workplaces, for instance multiple users share a screen but have their own mice and keyboards. Within the VE of these gamified IVR scenes, users are sharing a display since a team of them are interacting with the same work environment though their avatars. From the work of Beserra et al. (2014), users with more neighbours may experience greater cognitive load and less learning. With more complex IVR games this effect could be an interesting consideration designing the features of a team and distributing its members, but probably isn't for the lower complexity games that were available on the platform.

4.5 Evaluation of Platform

Based off the analysis of the major scene types and general characteristics of the platform, a general evaluation of the platform with respect to the recommendations of CLT can be made. Where the platform and its content met the criteria varies between the scene types, but there are many commonalities between them. Unfortunately, the opportunity to cross check the codes was missed, had it been done the researcher would have used the platform and applied the criteria, from which a comparison and discussion for congruency would have been made. How the content of the scene met the criteria has been discussed in detail. Overall the content that was available on the platform met the criteria well, recommended changes and general recommendations were discussed. Importantly, through analysing the content an analysis of the intrinsic characteristics and limitations of the platform, and how they relate to the reduction of extraneous load, was also made, from which the evaluation will follow.

Q1 The platform meets the criteria well for several reasons: The instructor directly participates in and controls the flow of content; The platform provides tools for the instructor to manipulate content and the flow of information; The platform (using GearVR) allows instructors and users to communicate normally; The platform allows for shared VR experiences with multiple students; The platform allows the instructor to create and adjust teams and monitor relevant data. By participating in and controlling content, the instructor observes the content as the users experience it, and can stop and change the content when necessary. The platform provides many ways to do this, an instructor can use the drawing tool to create visual signals and connect visual elements to improve spatial contiguity, they can pause and replay content to segment it, and other forms of control that are dependent on the scene such as zooming and rotating models. By allowing a classroom to communicate normally, the instructor can communicate with users to determine when to manipulate the content, such as pausing a video or pointing out critical information with the drawing tool. This also means that the instructor can vocally instruct like normal or provide audial cues. Many users can participate in the same content which helps the instructor keep track of the users making it easier to interact with them. Large numbers of users opens up room for collaboration, the platform can also support content in which users occupy collaborative team roles. The instructor can create and manage teams of users to ensure that they have the right composition and team experience. The instructor can monitor the progress and performance of teams of users using the data the system records, from which they can identify which teams need their attention. They can then guide students by through ordinary speech or creating visual signals with the drawing tool.

Q2 Several potential improvements, and considerations, were discovered from the analysis. When content is user paced, or an audial element is central to the intended learning outcomes, written text will be superior to spoken text. Instructors could be given the option to toggle subtitles during dynamic content. In user paced VEs with a lot of written text, pop-up text could be a viable option to not only spatially integrate text and declutter the VE, but to also add some user interaction. Creating more tools that instructors can use to manipulate any content, like the drawing tool, would give instructors more options to reduce extraneous load. In particular, tools that help users navigate to the important content would be a welcome addition. In a similar vein, user could be assisted by a navigation UI element, such as a compass, to help guide their attention towards important content without the assistance of the instructor. A final consideration relates to using the platform in class. As users can hear each other they could be distracted by unrelated sounds from the classroom. Teachers could consider classroom rules and standards for students to follow when the platform is used to reduce this, these could even be integrated with safety guidelines.

5 Conclusion & Recommendations

This study suggested that a solution for improving enrolment, retention and learning outcomes in STEM subjects was the correct implementation of IVR in the STEM classroom. It was identified that there is need to develop IVR with educational research in mind, for this study in particular this involved following the principles of CLT. This study attempted to apply CLT to evaluate an educational IVR platform by analysing how it compared to criteria synthesised from the principles of CLT (Q1), following with recommendations for possible improvements (Q2). The conclusion will discuss the outcome of a) the evaluation and b) the literature review, since interesting results regarding the state of the literature on the topic of IVR and IVR-CLT were discovered. These results determined some gaps in the literature, future research directions as well as recommendations for future researchers, all are discussed in the recommendations section following the conclusion.

5.1 Conclusion

The evaluation of the platform concluded that it had significant potential for reducing extraneous cognitive load during its application, even in a classroom. This was thanks to the unique set of characteristics that define it. This conclusion was determined after directly using platform by observing and using the content and characteristics/tools that were available. These observations were conducted using a coding framework consisting of criteria that was synthesised from the principles of CLT and the characteristics & limitations of the platform initially identified. Through these structured observations, extensive analysis and discussion regarding content design and minimising extraneous load within the VEs were made. More importantly, in these discussions the defining characteristics of the platform were refined and their interaction with the criteria was determined. From this, an evaluation of the platform against the recommendations of CLT (for minimising extraneous load) was completed. Finally, recommendations for improving the platform's capability to minimise extraneous cognitive load were suggested, also based off the analysis and discussion of the observations. As far as the author is aware, an attempt to evaluate an IVR platform against CLT has not made before. Hopefully, this study could help to guide or inspire future evaluations of ICT against educational theory, at least to some extent.

It was discovered that there existed numerous significant gaps in the literature, all of which guided the assumptions that had to be made when designing the evaluation. There is generally a paucity of evidence demonstrating correlations between IVR's affordances & implementation, pedagogy and learning outcomes. The most significant of which were the connection between CLT and: Immersion & Presence; Interaction & Interactivity; HMD usage. Furthermore, the connections between these factors and learning outcomes is also in need of further research. These findings are elaborated on in

5.2.1. in order to conduct the evaluation, several assumptions were made with respect to these factors. Furthermore, the evaluation itself relied upon a subjective approach. To increase validity the author’s background and biases were discussed and extensive descriptions were used in analysing the results. Unfortunately, the author missed the opportunity to assure the reliability of the study, but the approach was considered and planned out.

The inspiration for this study was adjusting our approach to STEM education to promote student interest in STEM subjects and improve their educational experience, in the interest of improving engagement, retention and learning outcomes. The literature leaves little doubt for the affective potential of IVR, its ability to motivate and foster interest in students stands out even among ICT technologies. IVR will clearly contribute to the drive for an embedded STEM curriculum, whose features were summarised in the introduction by Chen (2001), as artificially creating these experiences is precisely what defines it. But the drive to implement IVR into education must not be based upon intuitive assumptions of what it can do, but rather solid empirical research and educational theory regarding its effects on long term learning outcomes. Similarly, this study primarily focused on IVR alone as an educational tool, but in actuality the implications of IVR will only be understood once the technology has been deployed in real classrooms alongside traditional methods (Jensen & Konradsen, 2018; Lateef, 2010; Menin et al., 2018; Southgate et al., 2019).

5.2 Recommendations

5.2.1 Gaps in the Literature & Future Research Directions

The results of this study raise some ideas for future possible research. With respect to the platform itself, a major limitation of this study is that it did involve students and measure their learning outcomes and experienced cognitive load. A basic experiment could be to test the platform in an experiment with students, in secondary school, either by comparing it with a traditional or non-VR medium, or the current platform with an altered version guided by CLT (such as the design criteria of this study). These studies could verify how the platform compares to traditional approaches in an ecologically valid setting or whether the changes recommended by CLT produce positive learning outcomes for the platform. This study also focused on the scenes available, designing scenes from scratch with CLT in mind could further test the limits of the platform as opposed to simply altering existing scenes.

Regarding IVR in general, the most significant gap in the literature that was relevant to this study is the lack of a precise relationship between CLT and the concepts of immersion, presence and motivation. More generally the relationship between IVR and cognitive learning outcomes is in need of research (Lu et al., 2018; Makransky et al., 2019; Whitelock et al., 2000).

This raises numerous possible research directions. Makransky et al. found that the redundancy principle of CTML seems to translate well to IVR, but the other principles require further research to fully verify their applicability to IVR. They suggest that there might be situations in which added presence could increase learning and transfer, for instance it could be better employed for more advanced scientific learning such as realistic visualisations. Additionally, they suggest longitudinal research of courses which integrate IVR to determine whether the overall learning outcomes of the course are improved, particularly devices that can use a student's phone. This direction is especially relevant as even if IVR is used for its effect on affective outcomes, its use might increase the effort and interest of students such that they learn better in the non-IVR components of their classes. This approach suggests research based within more authentic environmental settings, also supported by Parong & Mayer, Jensen & Konradsen & Mikropoulos & Natsis. This is often called *Ecological Validity* (Olk et al., 2018), the extent to which a study approximates real life settings. Ecologically valid research will be important for IVR in education as things rarely work as planned within classrooms, however ecologically valid research presents issues of its own.

Research into the display and peripheral devices, primarily their effect on cognitive learning outcomes, would also be fruitful. As identified in the literature review, there appears to be mixed opinions on the relationship between these devices and learning outcomes. This includes HMDs alone, Jensen & Konradsen suggest that the question for future research is "not if HMDs should be used, but rather how and for what should HMDs be used" (2018, p. 1526).

For 360˚ video, several research directions are noted in the literature. Hendriks Vettehen et al. recommend research that compares the immersive characteristics of IVR with regular video as well as the characteristics of the messages presented in the video, such as their complexity. Message complexity is a decisive condition for several effects described in CLT, so examining the added effect of immersive displays will be important. Feng recommends that researchers study navigation within 360˚ video, as it was found that 360˚ video outperformed regular video when viewers could navigate smoothly and guided by visual elements like cues.

Creating a detailed set of empirically based guidelines for designing IVR content, which incorporate the wisdom of multiple pedagogies is an ultimate goal for this field. The lack of a set of guidelines has been voiced by numerous studies, as mentioned in the literature review. Such a set will allow future developers of educational IVR to pick up and follow it regardless of their professional background.

5.2.2 Recommendations for Future Investigators

For conducting future research into the topic of IVR and learning, including the incorporation of CLT, this study has some recommendations. Firstly, for measuring cognitive load, the most efficient measure for a large group will most likely be subjective ratings from questionnaires. Questionnaire based method methods have proven to be surprisingly consistent (Sweller et al., 2011), although they have not perfect they are far less intrusive more flexible to implement than other methods (secondary task performance, physiological measurements). Furthermore, given IVR's high attentional requirements and tendency to detached users from the real world, the other measurement methods could be difficult to implement in such a way that does not impact the validity of the research. Secondly, when designing an experiment to measure cognitive load and cognitive learning outcomes, it is important to ensure that the intrinsic load of the task is optimal for the research participants. If the intrinsic load of the task is too high the participants are likely to experience difficulty under any conditions, too low and the participants will have enough WM resources to deal with even high extraneous loads such that no meaningful difference will be found. Thirdly, an issue noticed in this study is that there is variation in the definitions of key terms within the IVR literature. Investigators must ensure that they consider which meaning they are adopting in order to accurately assess the literature and or define the intended constructs such that testing is valid, etc. A final note is to always incorporate the price, especially the hardware costs, and safety implications when assessing IVR devices, the latter of which is still not well understood.

6 References

Alsawaier, R. S. (2018). The Effect of Gamification on Motivation and Engagement. *International Journal of Information and Learning Technology, 35*(1), 56-79. doi:10.1108/IJILT-02-2017-0009

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing : a revision of Bloom's taxonomy of educational objectives* (Complete ed. ed.). New York: Longman.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195): Elsevier.

Australian Government. Productivity Commission. (2016). *Digital disruption: What do governments need to do?* (174037584X). Retrieved from Canberra:

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829. doi:10.1038/nrn1201

Baddeley, A. (2010). Working memory. *Current Biology, 20*(4), R136-R140. doi:10.1016/j.cub.2009.12.014

Baumeister, J., Seung Youb, S., Elsayed Jillian Dorrian, N. A. M., Webb, D. P., Walsh, J. A., Simon, T. M., . . . Thomas, B. H. (2017). Cognitive Cost of Using Augmented Reality Displays. *IEEE Transactions on Visualization and Computer Graphics, 23*(11), 2378-2388. doi:10.1109/TVCG.2017.2735098

Beserra, V., Nussbaum, M., Oteo, M., & Martin, R. (2014). Measuring cognitive load in practicing arithmetic using educational video games on a shared display. *Computers in Human Behavior, 41*, 351-356. doi:10.1016/j.chb.2014.10.016

Bork, F., Schnelzer, C., Eck, U., & Navab, N. (2018). Towards Efficient Visual Guidance in Limited Field-of-View Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics, 24*(11), 2983-2992. Retrieved from https://ieeexplore.ieee.org/document/8456525/. doi:10.1109/TVCG.2018.2868584

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal, 9*(2), 27-40.

Chen, M. (2001). *A potential limitation of embedded-teaching for formal learning.* Paper presented at the Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society.

Cooper, N., Milella, F., Pinto, C., Cant, I., White, M., & Meyer, G. (2018). The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment.(Research Article)(Report). *PLoS One, 13*(2), e0191846. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5794113/pdf/pone.0191846.pdf. doi:10.1371/journal.pone.0191846

Creswell, J. W. a. (2018). *Research design : qualitative, quantitative, and mixed methods approaches* (Fifth edition. ed.): SAGE Publications, Inc.

Davis, R. O. (2018). The impact of pedagogical agent gesturing in multimedia learning environments: A meta-analysis. *Educational Research Review, 24*, 193-209. doi:10.1016/j.edurev.2018.05.002

de Jong, T. (2010). Cognitive Load Theory, Educational Research, and Instructional Design: Some Food for Thought. *Instructional Science: An International Journal of the Learning Sciences, 38*(2), 105-134. doi:10.1007/s11251-009-9110-0

Dede, C. (2009). Immersive interfaces for engagement and learning. *science, 323*(5910), 66-69.

Denscombe, M. (2010). *Good Research Guide : For small-scale social research projects*. Berkshire: Berkshire: McGraw-Hill Education.

Ericsson, K. A., & Kintsch, W. (1995). Long-Term Working Memory. *Psychological Review, 102*(2), 211-245. doi:10.1037/0033-295X.102.2.211

Feng, Y. (2018). Facilitator or Inhibitor? The Use of 360-Degree Videos for Immersive Brand Storytelling. *Journal of Interactive Advertising, 18*(1), 28-42. doi:10.1080/15252019.2018.1446199

Freina, L., & Ott, M. (2015). A Literature Review on Immersive Virtual Reality in Education: State Of The Art and Perspectives. *eLearning & Software for Education*(1).

Geary, D. C. (2008). An Evolutionarily Informed Education Science. *Educational Psychologist, 43*(4), 179-195. doi:10.1080/00461520802392133

Hanson, K., & Shelton, B. E. (2008). Design and Development of Virtual Reality: Analysis of Challenges Faced by Educators. *Educational Technology & Society, 11*(1), 118-131.

Harrow, A. J. (1972). *A taxonomy of the psychomotor domain: A guide for developing behavioral objectives*: Addison-Wesley Longman Ltd.

Hendriks Vettehen, P., Wiltink, D., Huiskamp, M., Schaap, G., & Ketelaar, P. (2019). Taking the full view: How viewers respond to 360-degree video news. *Computers in Human Behavior, 91*, 24-32. doi:10.1016/j.chb.2018.09.018

Higgins, S. J. (2003). *Does ICT improve learning and teaching in schools?* : BERA, British Educational Research Association.

Ibrahim, M. (2012). Implications of Designing Instructional Video Using Cognitive Theory of Multimedia Learning. *Critical questions in education, 3*(2), 83-104.

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *The Official Journal of the IFIP Technical Committee on Education, 23*(4), 1515-1529. doi:10.1007/s10639-017-9676-0

Johnson-Glenberg, M. C. (2018). Immersive VR and Education: Embodied Design Principles That Include Gesture and Hand Controls.(virtual reality)(Report). *Frontiers in Robotics and AI, 5*. Retrieved from https://fjfsdata01prod.blob.core.windows.net/articles/files/375272/pubmed-zip/.versions/1/.package-entries/frobt-05-00081/frobt-05-00081.pdf?sv=2015-12-11&sr=b&sig=y%2Fk8lFqjFyAJ17z%2F5gaqj55Y8aCdu8TOFtMfneiKSbY%3D&se=2019-03-06T13%3A08%3A54Z&sp=r&rscd=attachment%3B%20filename%2A%3DUTF-8%27%27frobt-05-00081.pdf. doi:10.3389/frobt.2018.00081

Johnston, E., Olivas, G., Steele, P., Smith, C., & Bailey, L. (2018). Exploring Pedagogical Foundations of Existing Virtual Reality Educational Applications: A Content Analysis Study. *Journal of Educational Technology Systems, 46*(4), 414-439. doi:10.1177/0047239517745560

Josephsen, J. (2015). Cognitive Load Theory and Nursing Simulation: An Integrative Review. *Clinical Simulation in Nursing, 11*(5), 259-267. doi:10.1016/j.ecns.2015.02.004

Kalyuga, S. (2011). Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educational Psychology Review, 23*(1), 1-19. doi:10.1007/s10648-010-9150-7

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

Khalil, M. K., Paas, F., Johnson, T. E., & Payer, A. F. (2005). Design of interactive and dynamic anatomical visualizations: The implication of cognitive load theory. *Anatomical Record Part B: The New Anatomist, 286*(1), 15-20. doi:10.1002/ar.b.20078

Kirschner, F., Paas, F., & Kirschner, P. A. (2011). Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology, 25*(4), 615-624. doi:10.1002/acp.1730

Kirschner, P., Sweller, J., Kirschner, F., & Zambrano R, J. (2018). From Cognitive Load Theory to Collaborative Cognitive Load Theory. *An Official Publication of the International Society of the Learning Sciences, 13*(2), 213-233. doi:10.1007/s11412-018-9277-y

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). Taxonomy of educational objectives: Handbook II. *Affective domain. New York: David McKay*.

Krokos, E., Plaisant, C., & Varshney, A. (2018). Virtual memory palaces: immersion aids recall. *Virtual Reality*, 1-15.

Kyaw, B. M., Saxena, N., Posadzki, P., Vseteckova, J., Nikolaou, C. K., George, P. P., . . . Car, L. T. (2019). Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. *Journal Of Medical Internet Research, 21*(1). doi:10.2196/12959

Lang, A. (2006). The Limited Capacity Model of Mediated Message Processing. *Journal of Communication, 50*(1), 46-70. Retrieved from https://doi.org/10.1111/j.1460-2466.2000.tb02833.x. doi:10.1111/j.1460-2466.2000.tb02833.x

Lateef, F. (2010). Simulation-based learning: Just like the real thing. *Journal of Emergencies, Trauma, and Shock, 3*(4), 348-352. doi:10.4103/0974-2700.70743

Lee, S. H., Sergueeva, K., Catangui, M., & Kandaurova, M. (2017). Assessing Google Cardboard Virtual Reality as a Content Delivery System in Business Classrooms. *Journal of Education for Business, 92*(4), 153-160. doi:10.1080/08832323.2017.1308308

Lei, J., & Zhao, Y. (2007). Technology uses and student achievement: A longitudinal study. *Computers & Education, 49*(2), 284-296.

Leppink, J., van Gog, T., Paas, F., & Sweller, J. (2015). 18 Cognitive load theory: researching and planning teaching to maximise learning. *Researching medical education*, 207.

Lim, C. P., & Tay, L. Y. (2003). Information and communication technologies (ICT) in an elementary school: Students' engagement in higher order thinking. *Journal of Educational Multimedia and Hypermedia, 12*(4), 425-451.

Lu, C. M., Wu, P. L., Cheng, Y. M., & Lou, S. J. (2018). *Effects on patterns of learning-support design in immersive virtual reality system* (Vol. 9).

Madrigal, E., Prajapati, S., & Hernandez-Prera, J. C. (2016). Introducing a Virtual Reality Experience in Anatomic Pathology Education. *American Journal of Clinical Pathology, 146*(4), 462-468. doi:10.1093/ajcp/aqw133

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225-236. doi:10.1016/j.learninstruc.2017.12.007

Markowitz, D. M., Laha, R., Perone, B. P., Pea, R. D., & Bailenson, J. N. (2018). Immersive Virtual Reality Field Trips Facilitate Learning About Climate Change. *Frontiers in Psychology, 9*. doi:10.3389/fpsyg.2018.02364

May, P. (2004). Analysis of Computer Algebra System Tutorials Using Cognitive Load Theory. *International Journal for Technology in Mathematics Education, 11*(4), 117-138.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed. ed.). Cambridge

New York: Cambridge University Press.

Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning and Instruction, 29*, 171.

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist, 38*(1), 43-52. doi:10.1207/S15326985EP3801_6

Melo, M., Vasconcelos-Raposo, J., & Bessa, M. (2018). Presence and cybersickness in immersive content: Effects of content type, exposure time and gender. *Computers & Graphics, 71*, 159-165. doi:10.1016/j.cag.2017.11.007

Menin, A., Torchelsen, R., & Nedel, L. (2018). An Analysis of VR Technology Used in Immersive Simulations with a Serious Game Perspective. *IEEE Computer Graphics and Applications, 38*(2), 57-73. Retrieved from https://ieeexplore.ieee.org/document/8336833/. doi:10.1109/MCG.2018.021951633

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education, 56*(3), 769-780. doi:10.1016/j.compedu.2010.10.020

Miller, G. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review, 63*, 81.

Moesgaard, T., Fiss, J., Warming, C., Klubien, J., & Schoenau-Fog, H. (2015). *Implicit and explicit information mediation in a virtual reality museum installation and its effects on retention and learning outcomes.* Paper presented at the Proceedings of European Conference on Games Based Learning.

Moreno, R., & Mayer, R. E. (2002). Learning Science in Virtual Reality Multimedia Environments: Role of Methods and Media. *Journal of Educational Psychology, 94*(3), 598-610. doi:10.1037/0022-0663.94.3.598

Morrison, J. B., Tversky, B., & Betrancourt, M. (2000). *Animation: Does it facilitate learning.* Paper presented at the AAAI spring symposium on smart graphics.

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

Neuman, W. L. a. (2014). *Social research methods : qualitative and quantitative approaches* (Seventh edition, Pearson New International Edition. ed.): Pearson.

Newhouse, C. (2015). Measuring the meaningful use of ICT in schools: a learning environments attributes approach. *International Journal of Technology Enhanced Learning, 7*(4), 309-325.

Newhouse, C. (2017). STEM the Boredom: Engage Students in the Australian Curriculum Using ICT with Problem-Based Learning and Assessment. *Journal of Science Education and Technology, 26*(1), 44-57. doi:10.1007/s10956-016-9650-4

Olk, B., Dinu, A., Zielinski, D. J., & Kopper, R. (2018). Measuring visual search and distraction in immersive virtual reality. *Royal Society Open Science, 5*(5). doi:10.1098/rsos.172331

Paivio, A. (1986). *Mental representations a dual coding approach*. New York: Oxford University Press.

Paivio, A., & Lambert, W. (1981). Dual coding and bilingual memory. *Journal of verbal learning and verbal behavior, 20*(5), 532-539.

Panizzon, D., Corrigan, D., Forgasz, H., & Hopkins, S. (2015). Impending STEM shortages in Australia: Beware the 'smoke and mirrors'. *Procedia-Social and Behavioral Sciences, 167*, 70-74.

Parong, J., & Mayer, R. E. (2018). Learning Science in Immersive Virtual Reality. *Journal of Educational Psychology, 110*(6), 785-797. doi:10.1037/edu0000241

Passey, D., Rogers, C., Machell, J., McHugh, G., & Allaway, D. (2004). The motivational effect of ICT on pupils. *Department of Educational Research*.

Pintrich, P. R. (2003). Motivation and classroom learning. *Handbook of psychology*, 103-122.

Potkonjak, V., Gardner, M., Callaghan, V., Mattila, P., Guetl, C., Petrović, V. M., & Jovanović, K. (2016). Virtual laboratories for education in science, technology, and engineering: A review. *Computers & Education, 95*(C), 309-327. doi:10.1016/j.compedu.2016.02.002

Prinsley, R. T., & Baranyai, K. (2013). *STEM skills in the workforce: what do employers want?* : Office of the Chief Scientist.

Regan, E., & DeWitt, J. (2015). Attitudes, interest and factors influencing STEM enrolment behaviour: An overview of relevant literature. In *Understanding student participation and choice in science and technology education* (pp. 63-88): Springer.

Ritz, L. T., & Buss, A. R. (2016). A Framework for Aligning Instructional Design Strategies with Affordances of CAVE Immersive Virtual Reality Systems. *TechTrends: Linking Research and Practice to Improve Learning, 60*(6), 549-556. doi:10.1007/s11528-016-0085-9

Roberts, A. (2012). A justification for STEM education. *Technology and engineering teacher, 71*(8), 1-4.

Roberts, W. (2008). The use of cues in multimedia instructions in technology as a way to reduce cognitive load. In E. N. Wiebe (Ed.): ProQuest Dissertations Publishing.

Schrader, C., & Bastiaens, T. J. (2011). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior, 28*(2). doi:10.1016/j.chb.2011.11.011

Slater, M., Linakis, V., Usoh, M., & Kooper, R. (1999). *Immersion, Presence, and Performance in Virtual Environments: An Experiment with Tri-Dimensional Chess*.

Sorden, S. D. (2012). The cognitive theory of multimedia learning. *Handbook of educational theories, 1*.

Southgate, E., Smith, S. P., Cividino, C., Saxby, S., Kilham, J., Eather, G., . . . Bergin, C. (2019). Embedding immersive virtual reality in classrooms: Ethical, organisational and educational lessons in bridging research and practice. *International Journal of Child-Computer Interaction, 19*, 19-29. Retrieved from https://ac.els-cdn.com/S2212868918300151/1-s2.0-S2212868918300151-main.pdf?_tid=38f739a0-d158-44cb-aa11-ee8e739c4438&acdnat=1551433873_3b1dd5a286d4d9479ba1a202456581e5. doi:10.1016/j.ijcci.2018.10.002

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science, 12*(2), 257-285. doi:10.1207/s15516709cog1202_4

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory* (Vol. 1). New York, NY: Springer New York, New York, NY.

Taxén, G., & Naeve, A. (2002). A system for exploring open issues in VR-based education. *Computers & Graphics, 26*(4), 593-598. doi:10.1016/S0097-8493(02)00112-7

Tcha-Tokey, K., Christmann, O., Loup-Escande, E., Loup, G., & Richir, S. (2018). Towards a Model of User Experience in Immersive Virtual Environments. *Advances in Human-Computer Interaction, 2018*.

Thompson, M. M., Wang, A., Roy, D., & Klopfer, E. (2018). Authenticity, Interactivity, and Collaboration in VR Learning Games. *5*(133). Retrieved from https://www.frontiersin.org/article/10.3389/frobt.2018.00133

https://fjfsdata01prod.blob.core.windows.net/articles/files/327504/pubmed-zip/.versions/1/.package-entries/frobt-05-00133/frobt-05-00133.pdf?sv=2015-12-11&sr=b&sig=0DIutIgvzcK5h8KNvpagG%2FPQCOAbF%2BLVuzW%2F7wae0ks%3D&se=2019-03-08T13%3A41%3A47Z&sp=r&rscd=attachment%3B%20filename%2A%3DUTF-8%27%27frobt-05-00133.pdf. doi:10.3389/frobt.2018.00133

Tulving, E. (1989). Memory: Performance, knowledge, and experience. *European Journal of cognitive psychology, 1*(1), 3-26.

Ulrich, F., Helms, N. H., Frandsen, U. P., & Rafn, A. V. (2019). Learning effectiveness of 360° video: experiences from a controlled experiment in healthcare education. *Interactive Learning Environments*, 1-14. Retrieved from https://www.tandfonline.com/doi/full/10.1080/10494820.2019.1579234. doi:10.1080/10494820.2019.1579234

Van Merriënboer, J., & Ayres, P. (2007). Research on Cognitive Load Theory and its Design Implications for E-Learning.

Vergara, D., Rubio, M. P., & Lorenzo, M. (2017). On the Design of Virtual Reality Learning Environments in Engineering. *Multimodal Technologies and Interaction, 1*(2), 11. doi:10.3390/mti1020011

Voogt, J., Knezek, G., Cox, M., Knezek, D., & ten Brummelhuis, A. (2013). Under which conditions does ICT have a positive effect on teaching and learning? A Call to Action. *Journal of computer assisted learning, 29*(1), 4-14.

Webster, R. D. (2014). *Corrosion Prevention and Control Training in an Immersive Virtual Learning Environment.* University of Alabama at Birmingham, Graduate School,

Westby, C. (2018). Cognitive Load and Learning. *Word of Mouth, 29*(4), 8-12. doi:10.1177/1048395018759556b

Whitelock, D., Romano, D., Jelfs, A., & Brna, P. (2000). Perfect presence: What does this mean for the design of virtual learning environments? *The Official Journal of the IFIP Technical Committee on Education, 5*(4), 277-289. doi:10.1023/A:1012001523715

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence, 7*(3), 225-240. Retrieved from https://www.mitpressjournals.org/doi/10.1162/105474698565686. doi:10.1162/105474698565686

Wouters, P., Paas, F., & van Merrienboer, J. J. G. (2008). How to Optimize Learning from Animated Models: A Review of Guidelines Based on Cognitive Load. *Review of Educational Research, 78*(3), 645-675. doi:10.3102/0034654308320320

Xie, H., Wang, F., Hao, Y., Chen, J., An, J., Wang, Y., & Liu, H. (2017). The more total cognitive load is reduced by cues, the better retention and transfer of multimedia learning: A meta-analysis and two meta-regression analyses. *PLoS One, 12*(8), e0183884.

Zhang, P. (2008). Motivational affordances: Reasons for ICT design and use. *Communications of the ACM, 51*(11), 145-147.