Using Virtual Reality to Improve STEM Education

Hossain Samar Qorbani

Supervisors:

Dr. Ali Arya
Dr. Chris Joslin

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Ph.D.)

in

Information Technology

Carleton University
Ottawa, Canada

Abstract

The purpose of this dissertation is to investigate the effect of using Virtual Reality (VR) technology on the students' experience in science education, particularly for interaction, learning, and accessibility. Science, Technology, Engineering, and Math (STEM) education has special requirements such as lab-based activities and abstract concepts that complicate setting up the environment and learning process. These complications are increased due to the pandemic and other remote access requirements. VR has unique affordances that make it a promising solution but its use in this regard has not been properly investigated and there are many open research questions related to its effect on interaction, learning, and accessibility in STEM education. Focusing on these three aspects, we ran a series of quantitative and qualitative studies to find out if the use of VR in science labs leads to an increased level of learning, efficiency, and accuracy of the tasks (measured by pre-post knowledge tests and the in-app data collection system). The Immersive/head-mounted VR (IVR) was compared to Desktop VR (DVR) and 2D/text-based conditions. Results indicated a significant difference in some areas particularly related to post-knowledge score, spatial skills, and learnability between 2D and VR conditions. Task completion rate, efficiency, and accuracy also indicated a significant difference between IVR and DVR groups, showing IVR performing better. The qualitative evaluation included an analysis of students' experiences with the use of VR and their perspective on its application for education. The result indicated that in areas such as ease of use, learnability, engagement, and overall satisfaction they preferred VR treatment better compared to 2D/ Text as they found the use of VR to be beneficial to their learning. Our studies also showed the efficacy of software-based accessibility features that improved interaction and learning for wheelchair users. We concluded that the implementation of a virtual environment for STEM education requires careful considerations in the design and implementation to make it technically practical to run on mobile Head Mounted Displays (HMD), be relevant based on established learning theories, minimize the effect of cybersickness, and be accessible for a wider range of audience. The findings of this research helped us propose *ScienceVR*, a framework for the design and development of VR experiences for science education.

Acknowledgments

First and foremost, I am extremely grateful to my supervisors, Dr. Arya and Dr. Joslin for their invaluable advice, continuous support, and patience during my Ph.D. study. Their wealth of knowledge and expertise along with their generosity to share them encouraged me in my academic research and beyond.

I would also like to thank Dr. Abdinejad and Dr. Dalili for their collaboration and support in my study; and Dr. Teather and Dr. Ahmadi for their comments and for critiquing my results.
My gratitude also extends to Quentin Valembois, Stanislav Kyian, and Allysa Misola for their technical and logistic support.

My special thanks extend to the faculty and admin team at the school of Information Technology at Carleton University, for their kind help and support that have made my academic experience a great one.

Thank you, Dr. Jim Davies and Ludivine Blais, for the opportunity to collaborate on the current study of using VR in cognitive science.

And finally, I would like to thank my family for their support and understanding over the past few years that helped me complete my studies.

Table of Contents

Figures

Tables

Abbreviations

AR	Augmented Reality
CLT	Cognitive Load Theory
DCS	Data Collection System
DOF	Degrees of Freedom
DVR	Desktop Virtual Reality
EDP	Engineering Design Process
FOV	Field Of View
HMD	Head Mounted Display
HDRP	High-Definition Render Pipeline
IVR	Immersive Virtual Reality
STEM	Science, Technology, Engineering, and Mathematics
UDP	Universal Design Principles
URP	Universal Render Pipeline
VE	Virtual Environment
VR	Virtual Reality

Chapter 1. Introduction

1.1. Motivation

Learning and teaching Science, Technology, Engineering, and Mathematics (STEM) have their own unique set of difficulties among many subjects in higher education. This is in part due to the abstract nature of concepts in STEM disciplines and because of the range of issues such as cognitive load, required spatial thinking, and experiential and hands-on learning. Most STEM-related fields require laboratory courses that are essential to provide the opportunity for students to interact and gain experience using scientific equipment, expensive instruments, and dangerous materials available in science labs. To address some of these challenges and issues, different types of educational technologies have been used to make the process of STEM education easier and more effective both for students and educators.

Scientific concepts can benefit from proper visualization using new modeling and display technologies. Many subjects in STEM fields contain abstract concepts and learning and absorbing them can be difficult for many students due to the cognitive load, as well as the nature of some topics that may be perceived as dry or boring. If learners are not engaged, they are likely to feel detached from the course materials and traditional lectures commonly used in educational settings [1], [2]. Studies show that the integration of technology and hands-on experiences into classrooms increases the chance to create engaging learning activities for students with diverse learning styles and preferences [3], [4].

Several scientific topics in STEM education require a dedicated physical laboratory. These labs often contain toxic and dangerous materials which can present a high level of danger if not maintained and handled carefully. Maintaining such physical facilities and the safety considerations around any science laboratory can be very costly. However, despite the existence of such physical labs in many educational institutions, difficulties in accessing different locations due to reasons such as remote learning or the COVID-19 pandemic prevent students from utilizing these facilities. We all experienced the recent health and personal safety regulations around social distancing designed to prevent the spread of the virus during a pandemic. Interactive 3D simulations can offer safe, inexpensive, and on-demand/remote possibilities for educational activities.

Within the past few years, hardware, and software for 3D Virtual Environments (3DVE), a.k.a. Virtual Reality (VR)[1], have significantly advanced and are becoming more affordable. VR is increasingly used in games and training applications [5]. More recent applications using mobile Head-Mounted Display (HMD) devices include firefighter training [6], construction workers' safety training [7], and experimental/pilot projects such as working with machinery in education [8]. VR technology is highly interactive and depending on the type of delivery platform it can provide different levels of immersion to users and support 3D visualization of complex objects and environments.

These platforms can include HMD-based immersive[2] VR (IVR), desktop VR (DVR), or even mobile devices (MVR). VR can provide the possibility to build virtual spaces such as science laboratories. Building authentic simulated lab experiments can provide a complementary tool for students to exercise complex science experiments regardless of geographical location and distance. Adding audio, visual, and haptic feedback in VR can also help to create a multi-sensory, fully immersive experience in a virtual science lab.

The increasing application of VR in different training scenarios presents an opportunity to investigate the efficacy of this technology in STEM education where learners manipulated objects and "learn by doing".

Due to its affordances [9] such as a realistic simulation and visualization, high level of interactivity including natural body movement/hand gestures, flexibility and customization for accessibility and inclusion, and immersive presence [5],[9],[10], VR (particularly IVR) is considered a potentially suitable tool to create realistic simulations where users get the opportunity to conduct experiments in a safe inexpensive virtual environment (VE) [11], [12].

VR technology enables us to build virtual learning spaces (3D virtual learning environments, 3DVLE) including science laboratories, and to create authentic simulated lab experiments [13],[14],[15]. Some potential advantages of using VR learning spaces include stimulating students to build mental models and a strong impression of more complex, abstract concepts [16], gaining required motor skills and muscle memory in preparation for the actual real-world scenarios [17], increasing engagement, motivation, and attention [18], and embodied cognition [19].

---

[1] While many people in the industry use the term VR for devices with full visual immersion through head-mounted displays, a common trend is to use the terms 3DVE and VR interchangeably. In this document, we follow this trend.
[2] Note that all VR systems regardless of using HMD can result in immersion. But in this thesis, we only refer to HMD-based VR as Immersive VR.

Despite the affordances of VR to provide visualization and interactivity, the use of VR technology in STEM education is still problematic and under-investigated, and the technology has not achieved a mainstream position in this field. There is no clear evidence of the efficacy of VR in STEM education [20] and while accessibility is an issue in both physical and virtual environments, there is very limited research on how to achieve it in VR [10].

1.2. Problem Statement and Significance

This thesis aims to investigate the problem of using VR in STEM education, answering some of the main research questions, and suggesting a framework to efficiently use VR affordances to meet STEM education needs, in particular authenticity of content and interaction, learning, evaluation[3], and accessibility. In the context of this document, we define accessibility as "the design of products, devices, services, or environments for people who experience disabilities to remove barriers for access and use.[4]". This definition is related to the notion of equity within the context of Equity, Diversity, and Inclusion[5] which goes beyond removing barriers and ensures full and respectful participation [21].

As we will detail in the following chapters, many STEM-related fields share characteristics such as abstract and complex concepts, spatial thinking, interactive labs, and hands-on exploration [22],[23][24],[25]. Due to problems related to cost, complexity, and safety, using virtual tools has been seriously considered a solution in STEM education [20]. The review of existing literature (see Chapter 2) shows that despite the growing interest, the use of 3D Virtual Environments (3DVE, a.k.a. VR) in education faces many challenges and gaps that have limited them to most experimental projects rather than mainstream usage. Development and production-related issues, needing specialized equipment and VR headsets, and training requirements for both teachers and students are an example of such challenges. The following issues are among the major gaps in the current use of VR, especially in science education, as described in more detail in Chapter 2:

---

[3] While this study includes a data collection system for assessment purpose, the exact methods, algorithms for assessment and using process metrics in performance evaluation are beyond the scope of this thesis. A fellow Ph. D student (Nuket Nowlan) has collaborated in some of the studies described in this thesis and has used the data to investigate the assessment portion.

[4] https://accessibilitycanada.ca/aoda/definitions/

[5] https://www.nserc-crsng.gc.ca/NSERC-CRSNG/EDI-EDI

- The authenticity of content and interactions: Most VR learning spaces are generic and suitable for basic activities such as browsing an environment and speaking to others. They are not customized for the specific needs of STEM education, which we will discuss. Existing virtual labs and other experiential learning environments, on the other hand, lack 3D interaction and object manipulation in a realistic and customizable way and are not designed for IVR.
- Learning/Educational Foundation: Current VR learning spaces don't have a proper connection to established learning theories and adequate methods for skill assessment. Their evaluation is also mainly concentrated on usability instead of assessment of learning processes and outcomes and specific educational needs in different fields such as STEM.
- Accessibility: Despite advancements in VR technology, the design of VR systems has not included accessibility and inclusion as a primary factor and those subjects have either not been considered or have been after-thoughts [10]. Some examples include physical controllers, content, and user interface design for persons with disabilities. These are especially important for STEM education, where engaging in experiential activities is essential.
- Evaluation: The evaluation of VR learning tools has not been comprehensive and adequate, especially for immersive cases using more recent Head-Mounted Display (HMD) devices, and including the above three major aspects: interaction, learning, and accessibility.

To investigate the application and effectiveness of VR systems in STEM education (considering the existing gaps), our studies focus on a series of specific research questions that if answered, can offer insight into the use of VR in STEM. We use these insights to propose a novel framework called *ScienceVR*, aimed to establish guidelines and methods for designing inclusive VR solutions that can be used for building virtual science labs. Our last study focuses on a particular aspect of inclusion which is accessibility for different groups with different physical abilities[6].

---

[6] http://nda.ie/Publications/Attitudes/Appropriate-Terms-to-Use-about-Disability/

1.3. Research Approach

Several problems in applying VR in STEM education are highlighted in the literature [26]–[29] and discussed in further detail in the related work section of this document. Based on this literature review (Chapter 2) and a pilot study (Chapter 3), we were motivated to develop a VR framework for STEM education that can help solve problems categorized into three main areas of interactions, learning outcomes, and accessibility. Based on our literature review, we identified interaction, learning, and accessibility as three major areas to be investigated. We defined a set of specific research questions to be answered by developing VR experiments and conducting user studies:

1. What are the most relevant educational theories applicable to a VR framework, and how can we apply them in the design of a VR system in STEM education? This question was answered through a comprehensive literature review detailed in the next chapter. Although in the literature different theories are named, there was not a clear and applicable linkage between these theories and VR applications.
2. Are VR systems more effective in teaching science lab-based procedures compared to the current teaching method (2D / Video)? If so, how immersive and desktop VR compares in interactions, usability, and efficiency in completing assigned tasks?
3. Can using VR and authentic visualization (using ball-stick models) improve learning compared to two-dimensional (2D/Text-based) content?
4. Can flexible interaction type impact efficiency, accuracy, and learning outcome, if it enhances or hinders user experience (UX)?
5. Can software controls for vertical movement provide improved accessibility for wheelchair users with task efficiency and completion rate similar to non-wheelchair users?
6. Can the improved accessibility (if any) improve learning?
7. What are the specific guidelines for accessibility and general guidelines for designing VR experiences for STEM teaching and learning?

To answer the above questions our research was completed in four phases (Figure1):

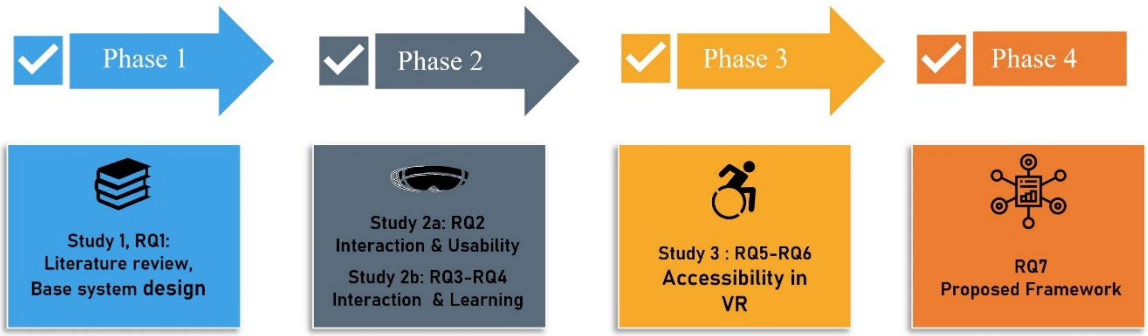


Figure 1.Research phases and corresponding Question (s) for each phase

Phase 1: focused on a literature review, learning theories, base system design, and Study 1 (the pilot). The main purpose of conducting a comprehensive literature review was to a) explore the status of technology-based STEM education, and b) examine how VR is used in educational settings. This study helped us to identify research gaps related to the use of VR in STEM education.

The pilot study was designed to evaluate the feasibility and practicality of a fully immersive virtual simulation of a science lab on a mobile HMD (i.e., Oculus Quest and Oculus quest2 seen in Figure2).

Figure 2.Oculus Quest headsets

OQ1 (left) vs OQ2 (right) with a faster processor and smaller form factor

Phase 2:

- Study 2a focused on general experience, primarily the interactions and usability while considering learning, data collection, and presence in both DVR and IVR. In study 2a we created a VR experience for both immersive VR using a Head-Mounted

Display (HMD) and a Desktop VR (DVR) version to compare them to the traditional two-dimensional (2D) method that uses text and video for teaching.

- Study 2b focused primarily on learning (the effect of Immersive VR and interaction techniques on learning). In this, we investigated and assessed the effect of using immersive VR on learning compared to text-based content. This was achieved by comparing a curriculum-based content delivery with two methods (Text vs IVR). We also compared two types of interactions within an immersive VR environment.

Phase 3: Most virtual environments offer the potential to reduce the limitations of a physical structure as they can be configured and use different control mechanisms. However, this potential has not been sufficiently investigated when it relates to accessibility. VR features such as "guardian" (i.e., a safe movement box around the user) may not be appropriate for those using a wheelchair who cannot maneuver in such a limited space (Figure 3).

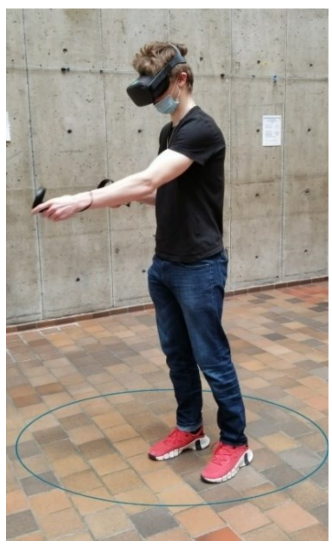


Figure 3.VR Safe area (Guardian)

The blue line (only visible in the headset) shows the “guardian”, a safe area, that can be set up within the headset to mark the border between the virtual and real world.

Therefore, study 3 was focused on VR accessibility (for wheelchair users). The aim was to investigate if using software controls can offer accessibility and learning advantages to wheelchair users in VR. By completing this study, we were able to compare and validate what kind of accessibility feature (s) could be developed to help wheelchair users access virtual objects in an IVR environment.

We concluded our research with a reflective process in Phase 4 that resulted in some guidelines integrated into our proposed framework, *ScienceVR*, which we used in the initial design of some VR-based actual course modules.

1.4. Contributions

In this thesis, we explored the efficacy of VR technology in science education with the following contributions:

- An insight into how affordances of VR can relate to the pedagogical needs of STEM education through our literature review and identifying theoretical framework answering RQ1.
- Understanding the effect of immersive VR and 3D interaction on learning in lab-based STEM activities. Our Studies 2a and 2b investigated this effect to address a gap in the literature concerning empirical evidence on the subject. The reported results are to answer RQ2, RQ3, and RQ4.
- A proposed method for improving the accessibility of elevation control in an immersive virtual environment answers RQ5 and RQ6.
- A proposed framework/ guideline for VR systems in STEM education that is authentic, interactive, immersive, accessible, and to some extent customizable answering RQ7.

The outcomes of these studies have been published (or are under review) in the following papers:

- Simulation and assessment of safety procedure[7] in an immersive virtual reality laboratory, Hossain Samar Qorbani, Ali Arya, Nuket Nowlan, Maryam Abdinejad. IEEE VR (Poster), 2021
- *ScienceVR*: A virtual reality framework for STEM education, simulation, and assessment, Hossain Samar Qorbani, Ali Arya, Nuket Nowlan, Maryam Abdinejad. IEEE AIVR. 2021
- Improving Accessibility of Elevation Control in an Immersive Virtual Environment, Hossain Samar Qorbani, Maryam Abdinejad, Ali Arya, Chris Joslin (Ready for submission)

[7] Safety procedure against the risks associated to the use of chemicals and accidents such as fire and chemical spills in a science lab

- Assessing Learning in an Immersive Virtual Reality space- A curriculum-based experiment in chemistry education (Work in progress for *Journal* of Science Education and Technology) Hossain Samar Qorbani, Ali Arya, Shadi Dalili, Maryam Abdinejad.

1.5. Thesis Outline

This thesis is structured as follows:

In chapter 1, we discuss the background, problem statement, research phases, and research questions.

Chapter 2 is dedicated to a comprehensive literature review, related work, theoretical framework, and gap analysis to answer RQ1.

In chapter 3, we discuss our research approach and the pilot study.

In chapters 4, 5, and 6, we detail the implementation and the results of each study conducted to answer RQ 2 to 6.

Finally, chapter 7 is dedicated to the proposed solution and some concluding remarks and directions for future research are discussed in Chapter 8.

Chapter 2. Related Work

In this chapter, we provide further details on the theoretical models seen as relevant to the use of VR including constructivism, cognitive load, scaffolding, experiential, and active learning theories. We also describe our search methodology, common characteristics, and needs of STEM education such as abstract concepts and the need for spatial skills to learn complex concepts. We also explain chemistry education as a STEM-related field offering plenty of lab-based activities that make it suitable for conducting VR experiments for academic research studies.

Further in this chapter, we describe VR affordances including interactivity and immersive experience as well as VR applications in education (in general and in STEM education), and the accessibility issues in VR. We conclude this chapter with a gap analysis, reflection, and a list of open research questions to be addressed by the subsequent studies detailed in the following chapters.

2.1. Search Method

Through these studies, our goal was to investigate the less-explored aspects of using virtual reality in STEM education and to provide a framework to efficiently use VR affordances that can meet STEM education needs in particular **learning, interaction, and accessibility**.

This study is a multidisciplinary research covering the fields of education, science (lab-based and cognitive), and information technology (Figure 4). These were the basis of our literature search.

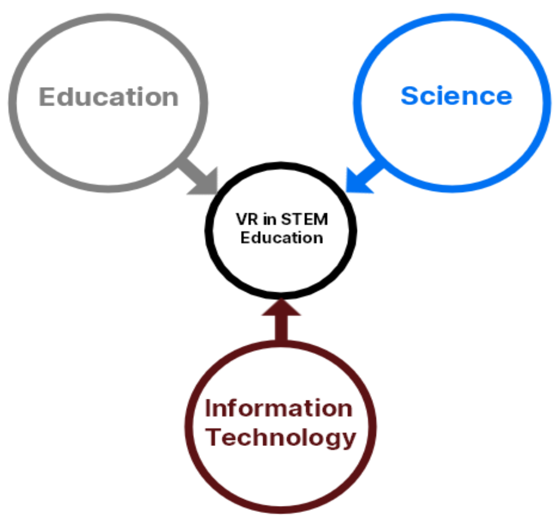


Figure 4.Multidisciplinary fields informing this research study

As we will detail in the next sections, modern forms of digital media content delivery platforms including AR and VR provide the possibilities for spatial navigation using body movement and hand gestures within the 3D virtual world.

Three major components of a digital experience are technology, people, and content. Considering the recent trends and gaps related to the use of VR, the virtual experiments we have built for our studies are laboratory-based science education using a certain area of information technology (i.e., VR) for undergraduate/college-level students as this group of audience is less studied as far as VR technology is concern ed.

To conduct a comprehensive literature review, we followed a process for academic research process described by Malcolm Tight [30] and methodological analysis of qualitative and mixed methods approaches by Snelson [31] that involve a systematic way of searching/importing articles, classification of sources, organizing internal and external sources, color coding the data and finally analyzing the data. Figure 5 shows the process we followed to conduct this literature review.

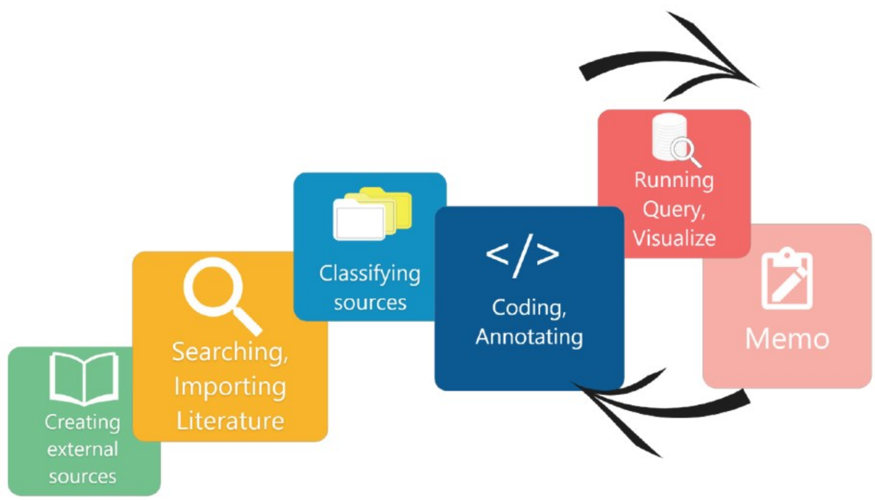


Figure 5.Academic research process

As the first step, full-text, peer-reviewed journal articles were selected and analyzed through a systematic review process. During this phase, relevant keywords and databases were selected based on a combination of established methods used in previous literature reviews [4]. Various known techniques for selecting manuscripts are mentioned in the literature including databases in which the studies are indexed; selecting all papers published within the leading journals of the field; and selecting a defined set of articles from the main journals within the field [32]. To limit the scope, we selected a defined set of papers from 2008 to 2019. While this document was developing several new and relevant articles were added in later stages. We only included published journal articles with the full online text and scholarly & peer reviews to manage the scope and rely on reputable studies. Adding a date range, specific discipline, and several other selections was used to narrow down the search. The keywords used for this pare included “learning theories”, “technology in the classroom”, “STEM education and technology”, “Augmented reality”, “Virtual reality in the classroom”, and “VR accessibility” as we wanted to gain a broad view of how technology is being applied in the classrooms to address STEM needs.

This research was conducted by one researcher based on a full-text search of the known databases including ACM digital library, Google Scholar, Summon by ProQuest, Web of Science, Carleton Library, and IGI as they were considered high-rank sources carrying verified, reliable studies. They were also cost-effective and accessible using Carleton student credentials to obtain full-text articles.

We updated our literature review during the consequent phases, to include more recent studies focused on VR, STEM education, and learning from 2013 to 2021. For this part, we searched and gathered references from multiple databases using the search phrase "AR/VR in STEM” and AR/VR in chemical education". We used SciFinder, Reaxys, Web of Science, and Scopus for this step.

After the completion of searches and importing all citations into the locally installed version of EndNote [33] the next step was to identify and remove duplicate citations. We used the “find and remove duplicate tool” in EndNote for this purpose. Then the citations from different databases were combined into a single group which resulted in two hundred ten (210) unique article citations.

We reviewed abstracts and full-text copies of the articles to determine eligibility for analysis. Articles were selected if they met the following criteria: (a) the study applied or discussed learning theories related to technology application in the classroom, (b) the study had a designed/developed AR/VR prototype, (c) the article was published in peer-reviewed conference or journal, (d) a full-text copy of the article in English was available. The results were categorized under three main categories:

- Learning theories
- Applications /Activities
- Accessibility

This process resulted in the final selection of one hundred sixty (160) full-text articles. However, further changes were made during the process as several articles were eliminated and several others added as new manuscripts became available.

In the next step, *NVivo, v.12 [34] software was used to code imported information. NVivo is a robust research tool that provides a centralized, single file that is easy to transfer and work with. It also creates an internal classification sheet that contains bibliographic information that is linked to each imported article. This classification sheet was used for running queries such as word frequency and text queries within NVivo or to export the results to MS Excel for further analysis when needed. In this phase, additional attributes were defined to create search/filtering terms to categorize each article based on the criteria emphasized in the study. By comparing these topics potential gaps could be found after running queries. A combination of word frequency and text search was used to help categorize and analyze the data.

2.2. STEM Education

2.2.1. Common Characteristics and Needs

A closer look at STEM education and lessons reveals that they are rooted in science experiences and a more hands-on or inquiry-based approach. We can categorize some of the specific characteristics of an ideal STEM education [35] as follows:

- Requires hard-to-set-up infrastructure, equipment, and safety procedures
- Content/subject matter is complex
- It may require extra skills and abilities (such as spatial skills) to be able to visualize and understand the concepts
- Involves active learning /learners' engagement
- Apply design methodology such as the Engineering Design Process (EDP)

Many STEM-related fields share the following common characteristics [36]:

- **Abstract concepts:** Many STEM lessons deal with abstract concepts that require extra efforts by educators to provide various learning tools and technologies to convey those abstract concepts to students. As a result, teachers may need extra training to use new technologies or learn new software to produce complementary information for their students. It is worth noticing that different STEM disciplines are dealing with the same issue.

For instance, in Physics there are concepts such as velocity vs torque, electric field vs magnetic field, and heat vs temperature that requires complementary illustrations and animations or extensive laboratory space and equipment to help students experiment and understand these concepts.

- **Spatial skills:** To translate textbook representations of abstract concepts into understandable ideas, many students may need additional skills, particularly spatial thinking to learn these complex concepts. Absorbing these concepts is difficult for students due to the cognitive load, as well as the dry and boring nature of some topics. Many cases in STEM education require considerable spatial thinking to the extent that some have argued spatial skills act as a gateway or are a barrier to entry into STEM fields. As explained by Uttal and Cohen [37] most of the research connecting spatial abilities and STEM education has focused on what Carroll [38] has referred to as "spatial visualization", which is the processes of apprehending, encoding, and mentally manipulating three-dimensional spatial forms". Relating two-dimensional representations to three-dimensional representations and vice versa are considered spatial visualization tasks. These tasks are relevant to thinking in many science disciplines including biology, chemistry, and physics.
- **Text-based content vs 2D vs 3D:** Similarly, subjects such as Chemistry deals with complex abstract concepts that make them extremely challenging to comprehend, particularly if educators only rely on their textbook and text-based content. Chemistry teaching also involves the presentation of numerous models using two-dimensional (2D) illustrations to help students explore abstract chemical concepts. These include microscopic molecular models and chemical structures (i.e., molecules or atoms) that are not visible to the naked eye. Studies also indicate that presenting these concepts in the 3D virtual form will help reduce cognitive load, therefore, enhance learning experiences in specific STEM-related fields.
- **Laboratory setup and space:** Many scientific concepts in STEM education require a dedicated laboratory within an existing school building. Several factors should be considered to build a laboratory including the location of the school concerning urban environment and accessibility, age of the building, available services, impact upon adjacent buildings and the scientific discipline to be taught in the classroom, whether that be for chemistry, physics, or others.

- **Chemistry**, as an example, includes laboratory experiments that may involve mixing chemical elements that create toxic fumes and reactions. This can present a high level of danger; therefore, strict cautionary and safety measures are mandatory to avoid accidents and harm to students. These safety considerations would make a science laboratory space very expensive to build and costly to maintain. Research studies suggest that despite the existence of safety standards and guidelines, accidents do happen in higher science education laboratories [39]. While most reported accidents were "relatively" minor, some have caused permanent injuries or fatalities [40].

- **Other characteristics:** In many STEM lessons activities/experiments are hands-on inquiry and exploration. These activities can engage students both individually and in teamwork. According to Winn et al. [41] body movement assists the learner to remember what they perceive and provides a cue for future recall of the information they have been exposed to in a lab environment. STEM education allows multiple approaches to finding the right answers. It also uses the Engineering Design Process (EDP), which includes defining a problem, conducting background research, ideation for a different possible solution, creating a prototype, then testing, evaluating, and redesigning. In many STEM subjects, course content uses 2D graphics and diagrams to help clarify complex concepts [17], [22], [23], [24], [25], [42], [43], [44], [45].

2.2.2. Chemistry Education

Chemistry is one of the STEM-related disciplines that follows the same downward trend of STEM education in decreasing numbers of graduates year after year [46]. As a complex field of study, Chemistry deals with many abstract concepts which are challenging without the use of analogies or models to improve students' conceptual understanding. In chemistry education, we can see all elements of abstract concepts, needed spatial skills, 2D/3D representation, physical models, and required laboratory experiments in teaching/learning processes.

Organic Chemistry, as a subdiscipline of chemistry, involves the study of structure, composition, reactions, properties of organic materials, and the different forms of carbon-containing compounds. Organic chemistry is very important in many industries including biotechnology, chemical applications, consumer products, petroleum, and pharmaceuticals.

Mastering this subject presents several challenges for students due to the abstract nature of the concepts that are often presented to them in flat, two-dimensional (2D) formats. The main barriers to learning Chemistry are the abstract nature of many of the concepts. As Gable described [47], chemistry instruction occurs in the most abstract level of its representations where mathematical symbols, formulas, and equations are used to represent complex concepts. Other levels of representation include the "macroscopic" level, which deals with laboratory experiments, and the "sub-microscopic" level, involving particle illustrations. To understand these concepts, students are expected to have/use a certain level of spatial skills.

Organic Chemistry, as a subdiscipline of chemistry, involves the study of structure, composition, reactions, properties of organic materials, and the different forms of carbon-containing compounds. Organic chemistry is a very important part of many industries including biotechnology, chemical applications, consumer products, petroleum, and pharmaceuticals. Mastering this subject presents several challenges for students due to the abstract nature of the concepts that are often presented to them in flat, two-dimensional (2D) formats. The main barriers to learning Chemistry are the abstract nature of many of the concepts. As Gable [47] described, chemistry instruction occurs in the most abstract level of its representations where mathematical symbols, formulas, and equations are used to represent complex concepts. Other levels of representation include the "macroscopic" level, which deals with laboratory experiments, and the "sub-microscopic" level, involving particle illustrations. To understand these concepts, students are expected to have/use a certain level of spatial skills.

This is also a challenge for instructors to successfully communicate these complex concepts because, in most cases, they are only using traditional methods from textbooks that have not changed in the last five decades. Spatial skills are essential to help students connect complex chemistry representations to conceptual and symbolic knowledge. As a result, when students are faced with complex representations of organic molecules, they begin to lose motivation to learn.

Students' conceptual knowledge of Chemistry is researched based on a constructivist model of learning in which students construct their concepts. According to this model (Figure 6), students build their own meaning from what is presented to them based on their background, attitude, and experience [46].

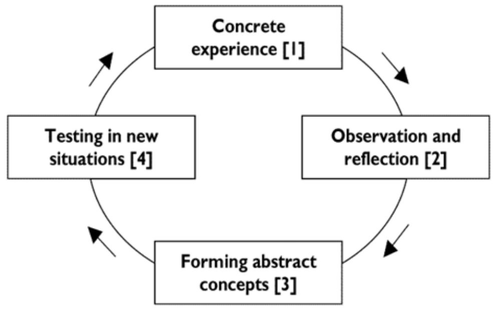


Figure 6.Kolb's cognitive model in Chemistry Education

Traditionally, 2D Figures or physical models, such as ball and stick molecular kits have been used to help students observe and build their concepts. However, studies show that using 3D models is more effective (compared to the 2D representation of chemical concepts. Stull et al. [48] proposed three ways in which 3D models, both physical and virtual, might help students develop "representational competence in chemistry. (1) 3D[8] models help off-load cognition as they reduce the load on working memory by what students must remember as they manipulate a physical or virtual 3D model. (2) Using multiple representations, models and drawings help them form more complete mental models. (3) 3D models help them connect new knowledge with prior and existing knowledge by "offering multiple perceptual modalities" meaning senses such as touch and 3D visuals should lead to better integration and absorption of new knowledge. Figure 7 shows this comparison of a 2D and 3D representation of a molecular structure [46].

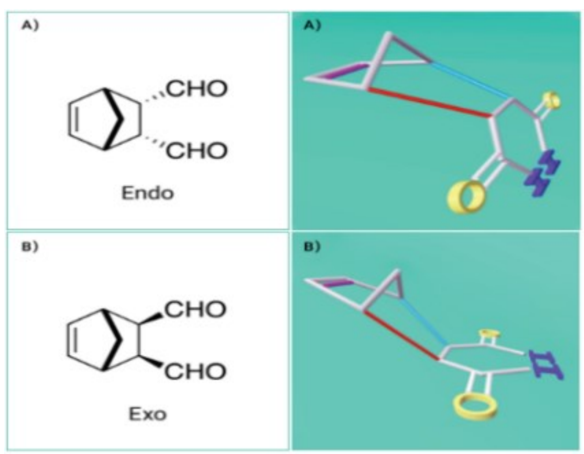


Figure 7.Comparing 2D (left) and 3D (right) molecular structures

[8] Wu and Shah [130] also mentioned that being able to comprehend chemical visualizations will be beneficial for students conducting advanced scientific research. In turn, these chemical visualizations become a common language which facilitates chemists or those who are interested in studying chemistry to communicate with each other (Kozma et al. [131].

2.3. VR Technology

The effectiveness of using 2D and 3D content in multimedia applications for science learning has been the subject of several studies [49], [50]. However, from 2000 to 2013 the research on using new technologies such as VR in undergraduate studies has been modest. The main issues related to these studies were the limited scope and small sample size and the experimental nature of the work [23].

Traditionally, the term Virtual Reality (VR) was used to refer to immersive fully virtual (computer-generated) systems using Head-Mounted Display (HMD) devices. It would be compared to Augmented Reality (AR) where the virtual content augments the real-world view. On the other hand, 3D Virtual Environment (3DVE) [51], was a more general term that included all 3D computer-generated spaces used in simulations, games, and other applications. A more recent trend is to use these two terms interchangeably [52] and divide them into platforms such as Immersive VR (IVR), Desktop VR (DVR), and Mobile VR (MVR), where only IVR uses HMD.

As suggested by Biocca and Delaney [53], VR is “the sum of hardware and software systems” that creates an all-inclusive, sensory illusion of being present in another environment. Three main characteristics of VR technology as described by Ryan et al. [54] are immersion, presence, and interactivity [53].

2.3.1. VR Affordances

General affordances described in the literature [55], [56] for VR, especially a virtual learning environment, include:

- **3D visualization and simulation:** VR (3D virtual environments) allows realistic or stylistic simulation of physical environments or visualization of non-existing imaginary ones. This not only makes hard-to-build environments like labs more available but also enables impossible experiences such as browsing through the human brain (See Chapter 7, Section 7.8).
- **Interactivity:** VR, especially immersive VR, provides a different way of interacting with content. Using handheld controllers (e.g., in Oculus/Meta Rift/Quest or HTC Vive) instead of a traditional keyboard and mouse, or a hand tracking sensor such as Leap motion sensor provides new ways to interact with digital content. This is particularly significant for simulating hands-on activities.

- **Engaging experience:** Depending on the type of VR explained earlier, the user may feel various degrees of immersion (full engagement), presence (being there), and embodiment (feeling a virtual body). These features are not the same, as a system may be immersive but not have embodiment or not offer a good sense of presence. Similarly, a non-immersive system may still provide a great sense of presence. But for the sake of this discussion, we group them under engagement/immersion, as investigation of presence and embodiment are beyond the scope of this thesis. One possible drawback in an immersive VR environment particularly with HMDs is that some people feel motion sickness while using it. Studies also suggest that "in an artificial learning environment, cognition is embodied in our physical action" (Winn et al. [41]).
- **Multimedia and multisensory:** Similar to other digital platforms VR displays show graphic representations such as diagrams, videos, and animations. The information can contain sounds and haptic feedback when manipulated while using sensors or the use of the body to interact with virtual objects.
- **Availability:** While not a technical affordance of VR, virtual reality hardware and software are becoming increasingly available. Although VR content needs relatively high computer capacity to process a large quantity of graphic information, the advancements in processing power make it possible to access smaller, faster, and lighter HMDs. Therefore, creating 3D content and animation for scientific concepts and making them available on these smaller platforms are becoming relatively easier every day [16]. As we will detail in the next section, VR content can facilitate conceptual learning for three reasons: (a) abstract concepts are presented in a dynamic 3D form that students can observe and explore, (b) students can interact, touch, and get haptic feedback from VR systems, (c) these interactions can stimulate students to build mental models and strong impressions of complex, abstract concepts.

Several design elements are used in building immersive VR experiences with visual, auditory, haptic, kinesthetic, vestibular, and olfactory components. Some of the design elements built on these components in an immersive VR include "basic interactions" to manipulate the virtual objects, "realistic environment", "immediate feedback" using haptic or visual feedback, "voice instructions", "traveling/teleportation" (moving around), "virtual rewards" and "knowledge test". As reviewed by Radianti et al. [53] most of the experimental studies done in VR applications have used a few of these elements with basic interaction and realistic surroundings being the most frequently used elements.

In our study, we apply and use all seven elements to create a fully immersive VR experience as explained in the next chapter of this document.

2.3.2. VR Applications in Education

Gutiérrez et al. [52] reviewed several studies from 1998 to 2014 to explain the advantages of virtual technology in education and how studies have shown that virtual technologies improved students' academic performance, social and collaborative, psychomotor, and cognitive skills and that virtual simulations help students to explore new domains. These were mostly desktop applications with limited interaction but showed the potential for using VR.

For example, the use of desktop-based virtual learning environments was illustrated in a series of courses by Arya et al. [13] involving the application of a virtual environment and automatic data collection to allow simulating hard to access settings such as excavation sites or downtown areas for prospective international students. They mentioned several advantages of such a virtual environment in education such as assessing higher-order skills such as critical thinking, communication, and collaboration to provide necessary support by instructors in the learning process.

In 2016, a review was conducted by Gutierrez et al. [52] to examine the "Virtual Technologies Trend in Education". Authors of this review intended to provide a "comprehensive understanding of AR/VR technologies" and discuss the possibility of using them in education.

They refer to the amount of investment made by big companies to develop headsets and platforms while emphasizing the fact that creating and developing content is necessary to make this technology popular.

Some useful applications named in this review included, virtual simulators to train professionals or maintenance in high-risk facilities such as nuclear power plants. Trends and future learning scenarios using AR/VR were discussed by first providing an overview of what AR/VR technology is and how it works. Types of head-mounted displays and hand/gesture tracking devices and external sensors in AR/VR were illustrated (Figure 8).

Figure 8.Popular HMDs for VR and AR

A more recent literature review by Scavarelli et al. [20] identified the primary environments in which VR technology was used to enhance learning. The examples provided included various AR/VR headsets, desktop, and mobile platforms, and 360 degrees videos. They identify content creation based on learning theories, accessibility, scalability, and technological limits as some of the existing challenges that require future research.

Based on these studies, the advantages of using AR/VR can be categorized as follows:

- Increasing learners' motivation and engagement and allowing a constructivist approach to learning. Students are active learners and are free to interact with virtual objects and other students; therefore, they get to build knowledge by investigating, conducting experiments, and getting feedback all resulting in an improved learning process.
- AR/VR technologies are affordable and relatively more available, and students can access shared VR content (mostly 360-degree videos) through online platforms such as YouTube.
- VR creates possibilities for students with certain disabilities to participate in experiences that would not have been possible otherwise.
- AR/VR allows more interaction and creates more immersive experiences.

Less focus has been observed on using VR for STEM subjects due to the difficulty of setting up the environments and experiences. We review some of the examples in the next section to explore the research gaps.

2.3.3. VR in STEM Education

While most use of game-based learning has been on non-STEM subjects [57], virtual environments in the form of 2D and 3D games have been used in STEM education for a while.

A study conducted by Mayo [58] in this area of using games in STEM education suggests the possibility to expand the reach of STEM education by using video games as the medium. By referring to the work done at Serious Games Summit, 2006, Mayo conveys that video games could yield a 40% positive learning increase over lecture programs and a variant of River City, and the ecology-related game had reduced the gap between grades D and B students and almost equalized it at the B level. The benefits named for the use of video games (vs lectures) include the just-in-time principle [59], which is the adaptability to the individual learning pace and style, as well as the use of multiple visual and auditory modes, the "multi-modal principle". In games, information and tasks can be presented gradually in smaller doses to let learners practice multiple times before getting to more complex activities all parts of what Gee calls the "incremental principle". Video games also provide immediate feedback as every keystroke produces a response through the gameplay. A system of rewards and points, progressive game levels, winning titles, or extra tools and capabilities provides motivation and can boost learners' self-confidence. It was also suggested that games invite more time on tasks, and she refers to the amount of time people spend time playing games per week. Teenagers generally spend 5-8 hours per week playing games which exceeds the time they spend on homework. Regarding the pedagogical and motivational elements in games vs lecture format, the author refers to earlier studies showing that switching to any interactive mode of instruction improved learning outcomes in the provided case of introductory physics [60].

A series of studies exploring the potential of using AR and VR in STEM [32],[53],[8], [61], [62] from 2013 to 2019 reaffirm the potential efficacy of VR for STEM education and how its unique affordances match the specific challenges of STEM. VR allows better visualization, hands-on learning, and improved engagement. "Atomic Structure" [63], Virtual Lab, and "Water Cycle in Nature" [64] were among the small-scale pilot studies that use virtual laboratory concepts and make a comparison between Desktop VR and Text/2D-based content. The authors point out the difficulties teachers must improve the motivation, engagement, and learning outcomes of students in STEM subjects. The lack of engagement is attributed to the perception that scientific subjects are difficult to learn. There is also a belief that interactive, engaging technology-based educational content can improve/increase learners' engagement [63], [64].

Another experimental project by Parong and Mayer [65] was conducted to compare the instructional effectiveness of immersive VR against a desktop slideshow (PowerPoint) for teaching scientific knowledge, as well as to examine the efficacy of adding a generative learning strategy to a VR lesson. The content of the experiment was a biology lesson about the human body that was delivered to a group of participants (55-57 college students) with two methods: 1) IVR and 2) PowerPoint slides. Authors expected that based on "interest theory", students who are learning the content on IVR will report a more positive rating on the "interest and motivation" in the post-test. However, for the students learning the content on the well-designed slideshow and based on the "cognitive theory", the expectation was that students would score higher on the post-test. An HTC Vive VR headset was used for IVR content delivery. The result confirmed that students who learned the lesson using IVR reported significantly higher ratings of motivation, interest, and engagement than the PowerPoint slide show but scored significantly worse on a post-test on the factual questions confirming the initial theories stated by the authors.

There has been serious interest and enthusiasm for use of innovative instructional technologies in Chemistry education. It is known that technology sessions at national meetings on Chemistry education are usually crowded and publishers provide technology support and content from simulation to visualization to online courses. Publishers are also interested in staying relevant and are exploring ways to combine new technologies, such as AR and VR, into their textbooks.

VR as an interactive media allows users to interact with virtual objects. In immersive VR (using HMDs), users are completely immersed[9] in a fully synthetic world which isolates them from the real one. AR/VR by nature attracts users' attention. Drawing students' attention is considered an important element of the learning process and educational technologies could be used to create a constructivist environment to enhance learning. AR/VR offers an alternative way (compared to traditional representations) to see the chemistry world which allows students to interact with the

---

[9] Virtual, Augmented, and Mixed reality technologies are described as:
• Immersive Virtual Reality (IVR) fully immerses users in a computer-generated artificial world that can be as Realistic as the real world (realistic level of details, texturing).
• Augmented reality (AR) overlays virtual objects (words, captions, animations, videos, labels) in the real-world environment.
• Mixed reality (XR) anchors virtual objects to the real world.

system as they discover knowledge on their own. AR/VR environments help to convey spatial cues directly to users as 3D models and visuals get presented in 3D space, meaning users would get a sense of spatial feeling. AR/VR is very useful in the knowledge domains that deal with spatial concepts [66], [67].

Other studies suggest that AR/VR can enhance learning as they allow users to directly interact with the content by using their body, especially their hands (as opposed to the traditional use of a keyboard and mouse), and they provide "sensorimotor feedback". This close and direct interaction helps users to manipulate the content, which in turn leads them to think carefully and be present and aware of their interactions. As a result, a better understanding happens due to paying more attention and focus [68].

On the other hand, these studies show that we need to create methods for assessment, and more polished, practical immersive content (in any given domain), that can be used by students. As identified in these studies most of the impressive VR applications done so far still need to run on a tethered VR HMD (Oculus/Meta Rift or HTC Vive) that is connected to a powerful computer. That is a limiting factor. Building authentic simulation and visualization of VR experience takes time and budget even more so if we aim to package it for mobile /untethered HMDs. Also, these applications use limited interaction, offer no or very limited data collection for assessment and guidance, and their pedagogical approaches are not directly based on learning theories.

The restrictions caused by the COVID-19 pandemic increased the interest in setting up virtual labs and other uses of VR in STEM education. A few companies now offer virtual solutions for general collaboration, training, and STEM education. Among them, we can mention the following products:

- Labster (https://www.labster.com): One of very few that actually has a STEM education focus and offers pre-made labs and some limited customization. Lack of customizable interactions, lack of under-grad level course content, and several other shortcomings make this solution an invalid option for our studies.
- LearnBrite (https://www2.learnbrite.com): A general-purpose avatar-based training VR. It is a costly solution for academic experiments.
- Virbela (https://www.virbela.com), Immersed (https://immersedvr.com), and Mozilla Hubs (https://hubs.mozilla.com): General-purpose VR environments that are used for various collaboration and presentation tasks. Lack of or restricted Application

Programming Interfaces (APIs) for data collection makes it difficult for the researchers to gain insights if they use this solution for academic studies.

More recent studies on the use of VR in STEM try to use these newly available 3D VR platforms for better interaction and use some well-known learning theories as the basis for designing VR experiences. For example, Zhao et al. [69] use a generative learning strategy and Lissalotte et al. [70] use self-explanation techniques to investigate the use of immersive VR in teaching biology. Their reliance on simple prototypes or existing limited VR products such as Labster[10] reduces their ability to have flexible interaction and freely customize the lab. These products also lack proper and customizable data collection that is essential in the assessment and guidance of students. But these examples demonstrate the potential for using customized VR experiences and motivate better educational theory-based experience design.

2.4. Accessibility and VR

A person with a special need is someone with limitations in body and structure functions, causing difficulties in participating in everyday life activities. Disability is a limitation or shortage of ability to operate an action in the way or within reach considered typical for a human being. It is not a description of a situation but an articulation of the person's connection to the surrounding environment [71]. The United Nations (UN) defines disability as a constraint or deficiency of the ability to carry out an activity as a normal healthy human being [72]. People with disabilities are the largest minority group, any person can join at any time, and due to accidents and old age, eventually do. Disability can be divided into eight main types of disability. Mobility/Physical, Spinal Cord (SCI), Head Injuries (TBI), Vision, Hearing, Cognitive/Learning, Psychological, and Invisible [73]. In 2017, one in five (22%) of the Canadian population aged 15 years and over – or about 6.2 million individuals – had one or more disabilities [74]. In this study, we concentrated on people with physical disabilities who use a wheelchair for mobility.

In the following sections, we discuss the interrelation of accessibility and VR technology in two aspects: using VR as assistive technology and making VR itself more accessible.

[10] https://www.labster.com/

Other aspects of inclusive design (such as gender, age, and cultural considerations) are beyond the scope of this thesis.

2.4.1. VR as Assistive Technology

In recent years VR is finding a place as an assistive technology for patients recovering from heart attacks to burn victims, phobia therapy, and post-traumatic stress disorder treatment [75],[76]. An important feature of VR is its flexibility which enables us to develop customizable applications to meet the unique needs of individuals with a disability. A limited number of studies have focused on physical disabilities using current VR controllers and other impairments including people with low vision. This particular study in 2019 by Zhao et al. [77] lead to the development of nine in-app tools for the Unity game engine to enhance the visibility of content for users with low vision. In another study by Maidenbaum and Amedi [78] the use of virtual assistive technologies such as "white cane via a haptic device" was proposed [79]. A virtual environment can be customizable to include or exclude certain features that would be helpful in each case. Such customization can be achieved depending on the goal of the application [80].

People with mobility impairments might require the use of a wheelchair while others may walk with the support of tools. The physical structure in most science laboratories is unwelcoming to people with physical disabilities and wheelchair users, or it is completely inaccessible. Many laboratories are difficult to navigate and visually obstructive for wheelchair users. In general science, labs are cluttered and full of obstacles, and not easily accessible for wheelchair users. Workbenches are high, cabinets are beyond reach and many equipment and tools are fragile [81].

As detailed by Sukhai et al. [79], "Faucets for sinks, gas hook-ups, power outlets, fume hoods and biological safety cabinets, eyewash stations, and other safety equipment also pose difficulties", since these areas of the lab are not easily accessible to wheelchair users [82]. On the other hand, the physical structures in most science laboratories are unwelcoming to people with physical disabilities and wheelchair users, or it is mostly inaccessible [83]. Doors, workbenches, sinks, fume hoods, and cabinets are examples of what cause accessibility issues related to vertical and horizontal movement to access different objects.

In this study, we focus on only a type of physical disability that requires the use of a wheelchair. Such use limits the user's vertical movement due to the constant sitting position they must assume during normal activities.

A useful characteristic of VR, as mentioned by Wilson, Foreman & Stanton [84], is that the need for semantics, symbols, or language in VR is virtually eliminated due to the experiential nature of the learning process, which allows many users from different ages and countries to enjoy using it.

VR can play the role of assistive technology as well. Lewis [85] hypothesized assistive technology as having two objectives: (1) to enhance a person's strengths to overcome the impacts of disability, and (2) another way to complete a mission that compensates for a disability.

Six key areas of accessibility to be considered by the Universal Design and VR community [86], [87] are defined:

- Accessibility of interaction techniques
- Accessibility of VR content
- Device/hardware accessibility
- Inclusive user representations within VR
- Environments
- Accessibility-focused applications for VR

And finally, SeeingVR [77] is a set of tools to make VR more accessible to people with low vision, and "Virtual Reality Without Vision"[88] is another recent experiment to make VR accessible to people with vision impairments.

Other studies related to people with physical disabilities were focused on handheld controllers and eye-tracking [89], [90] indicating that controlling devices should be simple and uncomplicated. There are limited studies to address wheelchair users' accessibility issues in VR. Therefore, we have focused on the accessibility of VR content, interaction, and its effects on performance and learning.

Individuals with physical disabilities face challenges daily [83]. Stairs are a common problem for those who have difficulty walking. Ramps and elevators are essential for moving about throughout the levels and the entire building for such individuals. They may face difficulty in accessing the classroom and finding suitable seating. Laboratory experiments or projects should be reassessed according to the limitations of the person with the disability.

Providing a personal assistant may be an appropriate way for students to participate fully in the curriculum [85], [91], Using VR technology as a distraction tool was proven to be an effective instrument to reduce pain for burn victims by making the patient concentrate more on the virtual world. Through such technology, people with special needs can focus more on their strengths, abilities, and learning preferences in comparison to the necessary learning task and projected learning outcome [92].

Wilson et al. [84] mentioned that in VR, the need for semantics, symbols, or language is virtually eliminated due to the experiential nature of the learning process, which allows many users of different ages and countries to enjoy using it.

Lewis [85] hypothesized assistive technology as having two objectives: 1) to enhance a person's strengths to overcome the impacts of disability, or 2) as another way to complete a mission that compensates for a disability. VR can act as an assistive technology in those two manners.

Users are expected to use modern VR handheld controllers to be able to manipulate objects in an immersive VR environment however as we are discussing the limited mobility and dexterity in hands and fingers other methods of interaction such as head tracking and eye-tracking come into play. In the area of eye tracking and head tracking in VR, a few studies were conducted to compare the eye-tracking and hand-held controllers for aiming. It was concluded that gaze could effectively replace hand controllers, and all tasks can be completed without a negative impact on the overall goals [89].

Several studies on the interactive control units [80],[90], [93] suggest two characteristics of preferable control devices for individuals with physical disabilities: 1) The controlling devices should be limited to two degrees of freedom, and 2) be simple, and uncomplicated.

2.4.2. Accessible VR

Another aspect of research on accessibility and VR is the focus on making VR technology and systems more accessible. To address accessibility as a high-priority issue, HMD manufacturers are beginning to include certain criteria for developers to embed accessibility features into their VR applications. The introduction of Virtual Reality Checks (VRC, https://developer.oculus.com/blog/introducing-the-accessibility-vrcs/) by Oculus in 2020 is one of such attempts focused on audio, visuals, interactions, movement, and other aspects of accessible design.

They are categorized into required and recommended VRCs to encourage developers to take serious effort and actions about making the software more accessible. Providing color blindness options, or the use of other techniques such as "color and pattern" combinations for easy visual detection is now a required VRC. However, there is no academic study on the application and efficacy of these techniques partially due to their novelty in VR.

On the subject of movement and motion controls, there are conflicting ideas and practices when it comes to user options. Currently, some users with physical impairments are excluded from playing some VR games due to the body position they are expected to take (such as standing). This excludes wheelchair- users. Some options are to adjust the body position or re-mapping the controllers to other controllers that users are familiar with [94]. The issue of a tethered headset (connected to a computer) is also mentioned frequently as a huge accessibility barrier to users with wheelchairs. The inability to customize the VR experience is one of the biggest issues raised in VR [95]. Involving people with disabilities in the design/development and testing process will enhance the accessibility of the final product. Some of the solutions offered by the industry to enhance the AR/VR experience (such as Relumino by Samsung for low vision users for example) either lag compared with other HMD development cycles, and require more expensive/powerful smartphones (i.e., Galaxy S8 and above) to run the AR/VR apps on, or do not work with other headsets due to the proprietary design [96].

Mott et al. from Microsoft Research [10] in their review of accessibility needs and opportunities in VR state: "Too often, the accessibility of technology to people with disabilities is an afterthought (if it is considered at all); or third-party patches to accessibility, while better than no solution, are less optimal than interface designs that consider ability-based concerns from the start. Virtual Reality (VR) technologies are at a crucial point of near-maturity, with emerging, but not yet widespread, commercialization; as such, VR technologies have an opportunity to integrate accessibility as a fundamental, developing cross-industry standards and guidelines to ensure high-quality, inclusive experiences that could revolutionize the power and reach of this medium." They identify content, interaction, devices, representations, and applications as major accessibility opportunities for VR [10].

Similarly, Menke et al. [86] reviewed the relationship between Universal Design for Learning and VR and concluded that the following research questions are still open:

- How can VR incorporate flexible training?
- How can content be presented?

- How can VR learning be personalized?

Our focus in this proposal is on the first two questions which are in line with the first two priorities stated by Mott et al. [10] while the third will also be considered.

Table 1 shows several correlations between STEM characteristics and the immersive VR features we intended to utilize in our proposed virtual science lab experience which can include, chemistry, physics, biology, and math.

Table 1. STEM Characteristics and VR Features

STEM education characteristics	Immersive VR features	Remarks
AREA 1- Infrastructure, equipment, and safety procedures • STEM lessons require physical laboratory space and exercises to enable learners to gain effective skills [22]. • Considering the increasing number of students in educational institutes, lab time is limited • Safety is a major consideration for science labs [23]. Safety measures, procedures, and setup makes these lab spaces costly to build, operate and maintain • Uses tangible objects to enable conducting scientific lab experiments	VR can provide the possibility to build virtual spaces such as science laboratories: • Authentic simulated lab experiments can provide a complementary tool for students to exercise complex science experiments regardless of geographical location, distance, or personal disability. • Using IVR as a complement to the real science lab can reduce the lab-time backlog • Students can repeat experiments without limitations in a safe and controlled environment • By adding Audio, visual and haptic feedback, VR can create a multi-sensory experience	Environments to compare: 1- Textbook, PowerPoint presentation 2-Desktop Virtual Reality (DVR).
AREA 2- Skills and abilities: • STEM disciplines require students to have • Spatial skills [42]. • Focus, Concentration, and attentiveness	Interactions in VR space can stimulate students to build mental models and a strong impression of complex, abstract concepts [16].	As explained by Uttal and Cohen [37] most of the research connecting spatial abilities and STEM education has focused on what Carroll [38] has referred to as "spatial visualization, which is the processes of apprehending, encoding, and mentally manipulating three-dimensional spatial forms".
AREA 3- Content/Subject matter: • STEM lessons focus on real-world issues and problems [24][43] • Course content uses 2D graphics and diagrams to help clarify complex concepts [44][97]. • Concepts are abstract, therefore difficult to understand (Cognitive load theory)[47]	In high-level applications from flight simulators to technical and medical training, VR has been used to create virtual simulations where users can be trained and gain required motor skills and muscle memory in preparation for the actual real-world scenarios [17]	The current mobile, low-cost VR headset allows us to begin building real-world experiences for Education. Our science lab is an attempt in this direction as we address the safety procedure simulation and visualization of scientific concepts.

AREA 4- Active learning /Learners' Engagement: • Activities/experiments are hands-on inquiry and open-ended exploration [43] • Engages students both individually and in *teamwork • Allows multiple approaches to find the right answers [24][25] *Our research does not cover collaborative VR experiences.	Students can interact, touch, and get haptic feedback from AR/VR systems and manipulate virtual objects. Research indicates that novelty, surprise, or uncertain events can attract students' attention and fascinate them [18]. AR/VR draws attention and is very engaging by nature as they present a new way of viewing and interacting with digital content. Studies also suggest that "in an artificial learning environment, cognition is embodied in our physical action" [19]. VR follows the inherent characteristics of "Digital Media" one of which is being programmable. Different scenarios and pathways can be programmed and embedded into the VR experience which provides alternative pathways to allow users to take a different approach to find the answer.	According to Winn et al. [41], body movement assists the learners to remember what they perceive and provides a cue for future recall of the information they have been exposed to in a virtual environment. We can use this feature to provide different paths to success in our virtual lab for safety procedures.
AREA 5- Design methodology and application: STEM lessons apply: • The Engineering Design Process (EDP) includes defining a problem, conducting background research, ideation for a different possible solution, creating a prototype, testing, evaluating, and redesigning[18][43]. • Math and science course promotes mathematics literacy[43]	In building any scientific virtual simulation, math and physics as gravity, force, velocity, etc. can be simulated. VR can be used to model, import, and build prototypes for engineering and scientific experiments	We use real-world math and physics rules to build a VR experience using the Unity game engine. For example in building a molecule structure, the correct angles between bonds and molecules and the right proportion of molecule size are just a few of the design consideration that follows math and science rules. Other rules such as gravity, speed, and acceleration can be also simulated in VR.

2.5. Educational Theories, Models, and Methods Applicable to This Study

In the previous section, we described that STEM-related fields such as Chemistry require complementary visual elements whether in the form of 2D, 3D or physical models plus laboratory space and equipment. These facilities enable students to gain experience and learn complex abstract scientific concepts. In this section, we explore the relationship between educational theories and the STEM fields covering only certain theories, methods, and skills that are relevant to AR/VR. The goal is to establish a link between those theories and the characteristics of AR/VR technology and gain insight into how we can apply them in the design and development of VR experiences for education, especially in STEM.

In the past few decades, learning science has made significant progress. There is a rich source of research-based theories that can explain how people learn. A set of evidence-based principles exists to show how to help people learn (based on cognitive theory). Several well-established educational theories, methods, and models support the use of AR/VR technology in STEM education. The assertion that AR can enhance learning experiences is grounded in two interdependent theoretical frameworks: situated learning theory [11] and constructivist learning theory [98], [99].

Current learning theories have been categorized by Merriam et al. [53] as constructivism, cognitivism, behaviorism, experientialism, connectivism, flow theory, and several others. However, we will only explain the ones that are most relevant to our study.

2.5.1. Constructivism

As a frequently cited learning theory in the literature, the constructivist theory was founded by Jean Piaget [53],[100], and it emphasizes the importance of the active involvement of learners in constructing knowledge for themselves [101]. This theory suggests that learning is an active, constructive process by which learners actively construct or create information from their subjective representation of objective reality. As explained by Sanders [26] "STEM education is grounded in the tenets of constructivism". He argued that STEM education pedagogy is, by nature, learner-centered and knowledge-centered.

[11] Situated learning theory explains that learning takes place within a specific context and the quality of the learning is a result of interactions among the people, places, objects, processes, and culture within and relative to that given context [98].

Von Glasersfeld [102] explains that constructivist learning theory proposes that students actively build new meaning by using their present conceptual framework to interpret new information that makes sense to them. They connect selected aspects of sensory input and existing concepts from long-term memory and integrate the newly constructed meaning. In a constructivist learning environment as proposed by Jonassen, learners are provided with problems while they try to solve them [103]. A high-fidelity immersive VR experience in this regard allows us to (1) present an immersive three-dimensional representation of a problem using multiple senses (visual, auditory, tactile, and/ or kinaesthetic) that mimics a real-world environment [104], and (2) present tasks, activities, and problems to allow learners try and solve them.

In a typical classroom, two types of activities occur: (a) what will the students do (the learning activities) and (b) what will the teacher do (the teaching activities). For many decades, even centuries, teachers have been presenting some organized summary of information from their understanding of the subject (lecturing) and leading occasional whole-class discussions using questions to get students' interest and to uncover new aspects of the subject. Such a conventional method creates learning experiences for students that lead them to listen, take notes, and occasionally participate in class discussions. In the last decade, this tradition has been challenged by the concept of active learning. As an alternative, Fink suggests five principles for delivering and creating a course to change this learning experience.

- A good course should challenge students to a significant kind of learning
- It should use active forms of learning
- Should have teachers who care (about the subject, their students, and teaching/learning)
- Should have teachers who interact well with students
- Should have a good system of feedback, assessment, and grading

2.5.2. Cognitivism and Cognitive Load Theory (CLT)

Cognitivism in psychology focuses on mental processes, including how individuals perceive, think, remember, learn, and solve problems. Cognitive theories emphasize making knowledge meaningful and helping learners organize and relate new information to existing knowledge in memory. It stresses the acquisition of knowledge and internal mental structures [105] and is concerned with how information is received, organized, stored, and retrieved by the mind. What learners *do* is not much of concern, but *what* they know and *how* they come to acquire it is important [106].

Cognitive Load Theory **(**CLT) is an instructional/learning theory that explains the cognitive process of learning. The assumption is that people's working memory has limited capacity. According to Baddeley's model (1986), working memory has at least two systems, a "visuospatial sketchpad" for processing visuospatial information, and a "phonological loop" to deal with phonological information. They have limited capacity and are independent. This limitation needs to be considered for instructional design/content [107]. A 3D virtual environment helps the learner to better visualize a problem presented to them in the context of the learning course in a rich format, therefore, reducing their cognitive load to build mental images and perform the activities.

2.5.3. Scaffolding

Scaffolding is the process by which teachers provide instructional support to students early in their learning and then gradually remove these supports to allow students to develop more knowledge and gain greater mastery [101]. It originated to describe one-to-one interactions with an ever-present tutor [108].

With the advancement of computer technologies, the following computer-based scaffoldings were designed to help students.

- Conceptual scaffolding, meaning what to consider when solving a problem
- Strategic scaffolding, building a strategy for addressing a problem
- Metacognitive scaffolding invites students to question their understanding
- Motivational scaffolding to enhance interest, autonomy, and self-sufficiency. Motivational scaffolding builds up the learner from within by using strategies such as showing concern, giving praise, and using optimism and empathy to encourage learners and reassure them of their capabilities.

Scaffolding enables students to become more proficient at skills such as problem-solving, argumentation, and evaluation. Studies show that scaffolding has consistently impacted positively on learning across student populations, STEM disciplines, and assessment levels. Scaffolding is now used in the context of many instructional methods such as project-based learning, problem-based learning, inquiry-based learning, and design-based learning [109]. An immersive VR environment can prevent a problem by adjusting the complexity. Scaffolding the learning process in VR can help learners develop skills, starting with lower fidelity simulation to direct learners' attention to essential elements and increase the fidelity and complexity as they progress [110].

2.5.4. Experiential and Active Learning

As defined by Kolb & Kolb [111] experientialism describes learning as a cycle of experiential stages from abstract conceptualization, active experimentation, concrete experience, and reflective observation. In essence, it is the process of learning through experience and learning through thinking and reflection on what learners do. Hands-on learning that is created in a VR environment can be a form of experiential learning. As described in the literature learning, is drawn from a learner's personal experience similar to constructivism [53],[4].

In a typical classroom, two types of activities occur: (a) what will the students do (the learning activities) and (b) what will the teacher do (the teaching activities). For many decades, even centuries, teachers have been presenting some organized summary of information from their understanding of the subject (lecturing) and leading occasional whole-class discussions using questions to get students' interest and to uncover new aspects of the subject. Such a conventional method creates learning experiences for students that lead them to listen, take notes, and occasionally participate in class discussions. In the last decade, this tradition has been challenged by the concept of active learning. As an alternative, Fink suggests five principles for delivering and creating a course to change this learning experience.

- A good course should challenge students to a significant kind of learning
- It should use active forms of learning
- Should have teachers who care (about the subject, their students, and teaching/learning)
- Should have teachers who interact well with students
- Should have a good system of feedback, assessment, and grading

According to Eison [112], active learning is, "anything that involves students in doing things and thinking about the things they are doing". Active learning strategies can be used to engage students in critical thinking and creative works, participation in speaking, (in small groups or with the entire class), creative writing and reflections, and giving and receiving feedback. However passive learning involves students only in receiving information and ideas while listening and taking notes from the lecture.

2.5.5. Guided learning, Segmentation, Signaling, and Gamification

As a proposed approach by Rogoff (1995), guided learning is defined as a process in which learners initiate and advance their learning guided by more experienced partners or mentors and socially derived sources such as tools, text, and/or other artifacts [113]. In the literature, "Instructional Strategy" refers to both the teaching method and the materials used in the teaching process. A test of learners' mastery of content is considered an assessment tool. However, as an instructional strategy, embedding tests in the content can positively impact the memory representation of the information and can enhance learning [114].

Signaling refers to the auditory or visual cues to assist learners to select and organize key information if a large pool of information is being conveyed [115].

Gamification is a term generally used to refer to the application of game mechanisms in a non-gaming environment to enhance the learning experience if used in an educational context [52].

In conventional text-based content, examples of signaling will include headings, outlines, arrows, colors, highlights, etc. In other forms of multimedia content, vocal emphasis, audio clips, animations, or hotspots, can be added to help learners focus their attention.

2.6. Gap Analysis

General trends and useful applications named in the literature included virtual simulators to train professionals or maintenance in high-risk facilities such as nuclear power plants [53]. Trends and learning scenarios using AR/VR in science education were named in the areas of Physics, Biology, Math, and electrical engineering [65] Several pilots and experimental studies developing applications in AR/VR (specifically in STEM education) were mentioned for different target audience groups from primary to secondary, K-12, college, and undergrad students, yet a very limited number of studies for students with special needs and physical impairments. Among the latest studies, only one study was for learners with hearing impairment [63] and one for children on the autism spectrum [116].

VR technology is being used in the areas of Physics, Biology, Math, and electrical engineering. Several studies in the application of AR/VR (specifically in STEM education) were mentioned for different target audience groups. In the following section, we provide a summary of the gaps identified in the literature.

The following list summarizes some of the major research gaps in the use of VR in STEM education:

- Despite the advancement and availability of AR/VR, the lack of a clear mapping from established learning theories to the learning goals and VR application development was identified as the first major gap.
- There is a lack of realism/fidelity in AR/VR content and experience, especially regarding STEM education and hands-on activities and interaction.
- There is a lack of curriculum-specific content.
- There are several attempts to see what AR/VR can do and achieve in education. The results suggest that AR/VR can motivate, attract, and interest learners; however, there are conflicting reports on the efficacy of this technology on learning, especially when it comes to IVR vs. DVR.
- The subject of inclusiveness and accessibility in VR is still a considerable gap in the literature, with three main areas of the flexibility of content and interaction, and various forms of personalization (which is related to content and interaction, but more specific to personal characteristics such as interests, needs, strength, age, and gender). Although everyone agrees that VR simulation can help users with accessibility/ mobility issues and special needs can benefit from this technology, related studies for this group of users are still very limited and not conclusive. There has not been a sufficient study on various ways of controlling movement and object manipulation, and also the presentation of content, to accommodate physical (dis)abilities.
- Proper assessment techniques in measuring the efficacy of AR/VR on the learning process remain another gap in the literature. This requires proper data collection and various algorithms to analyze and visualize the data. While this subject is beyond the scope of our work, we consider it part of the requirements for a VR framework for STEM education. The author of this thesis has collaborated with another PhD student who focused on this subject.

Overall, existing research studies show VR application in the context of STEM education has good potential to be utilized in a variety of educational learning activities. However existing work in this area is mostly limited to experimental development and usability testing with limited scope and scales.

Chapter 3. VR-based STEM Education

Our literature review helped us identify a series of research gaps related to interaction, learning, and accessibility in the use of VR for STEM education. It motivated us to develop a framework that offers guidelines and design considerations to develop VR experiences that are effective in those three areas. To investigate the possibility of such a VR-based STEM framework, we developed a base virtual environment and conducted a pilot study. This study showed the potential for using VR and the possibility of defining some design guidelines. It also confirmed specific research questions that needed further studies before such a framework could be established. In this chapter, we review the base environment and the pilot study setting up the roadmap for three main studies and the proposed framework.

3.1. Base Virtual Environment

Based on a set of specific educational theories mentioned in Chapter 2, Section 2.5 a realistic virtual environment (Figure 9) was built to enable learners to feel that they are in a similar physical environment and could try and solve problems. The base environment was used to conduct a pilot study to further evaluate the project's feasibility. It also helped us establish the main technical and theoretical features and was a starting point for different variations of the consequent studies (i.e., 2 and 3). Each variation is explained in its respective chapters.

This virtual environment provided an alternative to the physical lab to make distance learning possible particularly during the events such as the current pandemic. This virtual space can be further developed to use modular components to make it even more customizable to simulate any physical lab. Modularity can include adding customizable dimensions and floor maps to build a specific lab. By using a modern game engine to develop this framework, we can export the output for multi-platform applications including HMDs and Desktop computers. Embedding specific accessibility features for certain users with physical disabilities enables a wider range of audiences to use this framework.

Figure 9.Virtual chemistry lab as the base environment

3.1.1. Interactions

Interaction is defined as actions and operations that help users to manipulate, explore and understand. As the general assumption (by device manufacturers) is that all users are capable to make use of both hands and fingers to manipulate virtual artifacts, special consideration should be given to the types of interactions and tasks designed and implemented in virtual environments. Using a hand avatar, we can create touch, grasp, move, press, and drop interactions for the Oculus touch controller (Figure 10) to interact with 3D objects.

a) touch controller is visible b) touch controller is not visible

Figure 10.Oculus Touch controller + Hand Avatar

Interactions with objects include selection, grasping, and navigating in IVR using Oculus handheld controllers. To minimize the complexity of interactions the main roles of grasping

objects and teleporting were assigned to two main buttons on both controllers, i.e., the grip button and the trigger button (Figure 11). Since it is activated on both controllers' users can easily interact with objects regardless of their handedness.
In general, tactile controllers help users to get a better sense of touch when interacting with virtual objects. As controllers provide a key-press action and vibrate when a collision of the hand avatar with a 3D object occurs therefore providing better feedback to users compared to bare-hand[12] interactions. A better touch sensory feedback can happen when using haptic gloves that enable users to feel the virtual objects. Tactglove by Bhaptics is an example of this technology[13] that can be used in future studies of this nature.

- **Left joystick** _ To move around/ also helps to make smaller moves- May create dizziness for some users- **Right Joystick**: turns the head/view in 30 degrees increments
- **X button**- Opens a control menu
- **Y and B buttons**: Point toward an object if it is dropped on the floor, press to grab it remotely.
- **Oculus Home button**: Opens the oculus home menu
- **Grip button:** Press and hold down to grab objects that are closed to hands
- **Trigger button:** Point towards a target area then /press and release to teleport

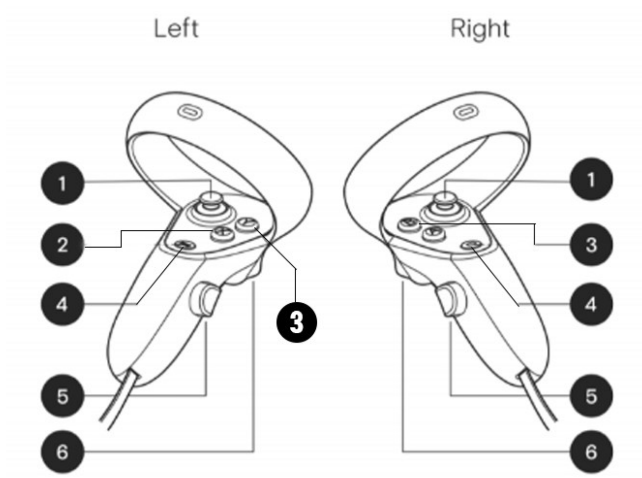


Figure 11.Oculus Touch controller and button functions

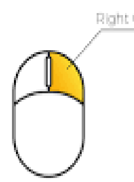
[12] Modern VR headsets allow hand-tracking as an option therefore users can interact with objects without tactile controllers
[13] https://www.bhaptics.com/

The same interactions on desktop VR use a combination of keyboard strokes and mouse clicks as shown in the following steps.

1) To move the character around, use WSAD or the arrow keys (Figure12)

Figure 12.WASD keys on a standard keyboard

2) To rotate the camera view (the character) hold the right mouse button and move the mouse cursor.



3) To select an object, point at it with the mouse cursor and click on it to grab, and collect the coins, use buttons in menus, and open the shelves (click on the shelve handle or door handle for example).

4) To release an object, press the left mouse button once.

5) To move the object, move the mouse cursor (for left, right, top, bottom) and use W and S keys on the keyboard (forward and backward movements).

6) To move the character with the grabbed object, hold the right mouse button (then WSAD or arrow keys can be used for moving, an object is moved along with the character, keeping the same coordinates relative to the character).

7) To rotate the camera with the grabbed object user, hold the right mouse button and move the mouse cursor (same as without the object). An object is moved along with the character view, keeping the same coordinates relative to the character view.

8) To rotate the object, hold the Left Ctrl button and use the WSADEQ keys to rotate the object around different axes.

Based on active learning and experiential learning theories, and by creating the 3D models of the equipment, we can replicate real physical lab equipment and mimic real experiences allowing learners to try things out (Figure 13). Working with equipment will allow students to gain experience operating them and gain motor skills and muscle memory to prepare for the actual process. Manipulating objects in the virtual world helps learners with spatial skills and understanding abstract concepts in a more effective way compared the 2D/text-based presentation of the same concepts, due to the hands-on inquiry and engaging exploration.

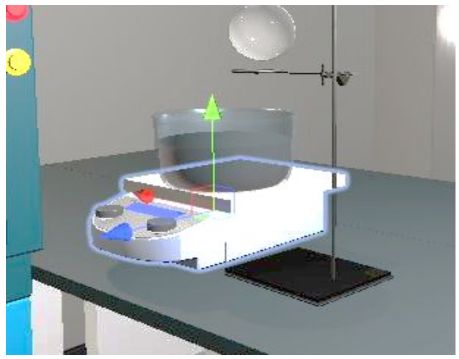


Figure 13.A hotplate used in a chemistry experiment

It is worth mentioning that although 3rd party solutions, such as Labster, exist for chemistry simulation, at the time of our study they did not include university-level environments similar to what we created for this our studies, which is specific and curriculum-based with flexible interaction. The user experiences in our case are tailored to a specific group of users and improved over 3 iterations.

Due to the time limitation and scope of the work, all prototypes and environments we have built were focused on single-user interactions. We hope to add multiplayer futures in future iterations and studies.

3.1.2. Ray Casting

Ray casting is a feature that is available for the selected HMD. For the immersive VR experiment in this study, we have built the interactions using both the handheld controller's feature and ray casting. Ray casting works by pointing a laser beam to a 3D object and holding a button/trigger on an Oculus touch device to hold/move/drop objects in VR space (Figure 14).

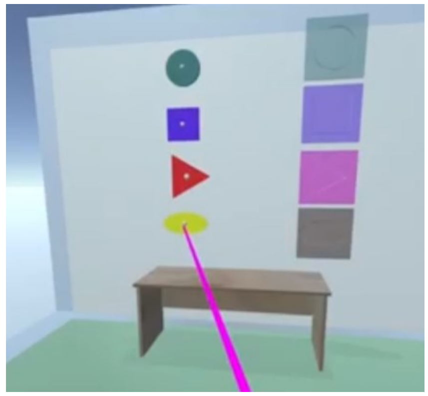


Figure 14.Ray-casting using handheld controllers

Raycast in Unity is a Physics function that projects a Ray into the scene. It returns a Boolean value upon hitting the target (an object in the environment). As a result, the information about the collision, such as the distance, position, or a reference to the object's change of position, can be stored in a variable for further use.

3.1.3. Other Design Elements

Other design elements that are included in this framework are sounds (the voice instruction, sound of knobs and clicks, fire and water), texts information (the help popup tips that are hidden under the question mark icon and are revealed when users point to them with the controllers), visual outline and highlights around 3D objects, haptic feedback (a short vibration in the handheld controller when users point to or grasp an interactable object) which will help decrease memory load (based on cognitive load theory).

Virtual reward (in the form of coin collection) for the task of finding/locating and finally an in-app quiz and in-app data collection system (see next section) to collect remote data for assessment (detailed in the next section). It also includes a feature that allows users on a mobility scooter to change the "player's" height. This is a proof-of-concept feature that was developed with accessibility issues in mind.

3.1.4. Data Collection System (DCS)

Due to the reasons mentioned earlier (noticeably, the lack of customizable data collection APIs), we created features based on object collision that allowed us to track and record users' interactions with objects in VR. The main goal for collecting such data was to be able to generate usable data for assessment since observing individual learners performing a task is extremely time-consuming. This assessment method relies on what is known as computer-based assessment to capture rich information on the learning process detailed in the work of a fellow Ph.D. student and collaborator of the study 2a, Nuket Nowlan. While our design includes data collection, the actual data analysis is beyond the scope of this thesis.

The data collector gathers information about the player's behavior. All interactable objects of the scene trigger a call to the data collector with the following elements:

- Interaction Type:

The interaction is of two types, Hover or Select. Hover happens when the players move their right or left hand near an object or when a ray is pointing toward an interactable object. Select happens when the player hovers over an object and engages in interaction with it by pressing a button. This can be to grab an object or to click on a User Interface (UI) element for example.

- Data Structure:

For each interaction, a call is made to the data collector that will write a line in the Log File at the end of the play session with the corresponding element i.e., TIMESTAMPS, NAME, USE TYPE, USED_BY, LENGTH

- **Timestamps:** The time the interaction took place
- **Name:** The name of the object that the player is interacting with
- **Use Type:** Either Hover or Select depending on the interaction type (see above for more information)
- **Used by:** The interactor used, either Left or Right hand if the player is interacting directly with the object or left or right ray if the player is interacting with the object using the Ray
- **Length:** The duration of the interaction between when the player starts the interaction and the end of the interaction. It is noteworthy that when a player "selects" a UI this value is 0 as there is no duration to that interaction.

- **Unwanted Behaviour:** To prevent from collecting unwanted behaviour from the player, Such as the player hovering over an object that he didn't intend to, which can be the case for example if he moves his hand toward an object that may be on the way to grab the object he hovers another one.

A time threshold is applied before collecting the data. If the *Length* parameter (see above for more information) is above this threshold the data is collected.

In the case of the use type "Hover," the threshold is set to 0.5. The same value is applied to the threshold when grasping an object. This means that to collect this data by the data collector, the object must be hovered for at least 0.5 seconds, and it needs to be grabbed for at least 0.5 seconds to have the interaction collected.

- **Player ID**

The data collector also gives the player a unique ID using the exact date and a random three-letter word. This allows us to have a unique identifier for each player and know the exact date of the play session. The gathered data is then sent to the researchers automatically by mail and saved locally on the device.

The data collection system was implemented in each of the built VR environments to record objective data related to task efficiency and completion rate. This system would record the type and duration of each interaction initiated by users while interacting with virtual objects and the environment. For task efficiency and accuracy, we used the logged data generated by the system, which included:

- Player's position and orientation (including body, head, and hand)
- Session start/finish time and calculated the duration
- Timestamp and duration of each interaction with objects both using raycaster and/or hands.

This feature eliminated any potential errors in data collection and did not interrupt the user testing process as participants did not experience any intermittent distraction to answer a question during the experiment. The collected data was automatically sent to a dedicated email address made for this study when participants completed the experience.

3.1.5. Specific Learning Theories Applied in Base VE Design

To answer RQ1, as discussed in the previous chapter (see Chapter 2, Section 2.6), the lack of a clear mapping and references to established learning theories for VR development was identified as a major gap. While different, scattered theories have been applied in separate studies involving VR implementations, we list the most relevant theories that can be applied to VR experiences and have an explicit link to the implementation of VR:

- Constructivism
- Cognitivism
- Scaffolding
- Experiential and active learning
- Segmentation
- Signaling
- Gamification

All these theories have been applied in the implementation of our sample VR environment:

- Based on Constructivism theory (see Chapter 2, Section 2.5.1) we created a high-fidelity immersive VR experience in which we can mimic real-world problems and allow learners to try and solve them. Simple problems such as stacking cubes or finding objects and operating equipment are examples used within our prototype environment.
- We applied the Cognitivism theory (see Chapter 2, Section 2.5.2), in this VR environment to only present the necessary information when invoked, and users have options to reveal or hide the extra information. In the virtual lab, we limited the number of displayed interactable objects such as personal protective or other lab equipment to only what a learner would need to see at a particular time during the training or testing phase (rather than adding all the possible equipment similar to what they would see in a physical lab).
- Scaffolding theory (see Chapter 2, Section 2.5.3) was applied to help users build skills using new VR controllers to interact with objects and find their way around the IVR environment. Starting with lower fidelity simulation to direct learners' attention to the most important elements and increase the fidelity and complexity as they progress. This will help to reduce cognitive load, particularly for first-time VR users.

- Based on Experiential and Active Learning theories (see Chapter 2, Section 2.5.4) our VR environment allows users to experiment with every interactable object and observe different possibilities without limitations. Related to EL, active learning is, "anything that involves students in doing things and thinking about the things they are doing" [112]. We used active learning principles in our VR experience to allow object manipulation and help learners to explore the environment and work freely and actively with objects, provide some feedback, and collect data for assessment.
- We used the Segmentation principle in our VR experience to categorize and break the information into smaller segments. In our virtual labs, activities are divided and presented in separate stations (desks) where users can walk or travel to complete a task. Safety procedures and related information are also broken into smaller texts and are hidden in the popup/dialogue boxes and are revealed only when users point/hover over them.
- And finally, based on the signaling and gamification principles in learning (see Chapter 2, Section 2.5.5.), we applied voice, a system of in-app quizzes and tests, and highlighting features for interactable elements corresponding to the learning strategies in this section. A system of rewards (coin collection) was applied to create a gamified vibe in the finding objects task.

3.2. Pilot Study

At the start of the project, a realistic virtual environment was built that functioned as a prototype for the pilot study. In this section, we first describe the design goals, participants, and procedure then we present the results and insights gained from the pilot study, and finally discuss how it informed future studies and the subsequent variations. The result of this pilot project was published at the IEEE VR conference 2021 as a poster with the title "Simulation and assessment of safety procedure in an immersive virtual reality laboratory".

3.2.1. Overview

As shown in the base virtual environment in the previous section, a prototype was built and tested on a limited number of university chemistry students. The goal was to evaluate the feasibility and practicality of using a mobile Head Mounted Display (HMD) for future studies. A safety training procedure and two accident scenarios were added to this virtual environment to help us study users' interactions with the environment and record their responses and reactions (see details in chapter 4). Safety training is a process each student should complete before entering a science lab that involves hazardous and toxic chemical materials.

In real-world educational scenarios, a standard training routine is provided to the students through reading materials and a standardized test known as Workplace Hazardous Information System (WHIMS[14]). Two accident scenarios added to this VR space are a small fire and a chemical spill detailed in the next section which allowed us to compare the effect of using VR in training and compare it to other platforms such as text (2D) and Desktop VR (DVR). Six participants volunteered to test the prototype. Figure 15 shows the process of user testing during our pilot study.

This VR environment and the safety training procedure formed the core of studies 2a and 2b which are discussed in detail in chapters 4 and 5 including the tasks and procedures completed by participants.

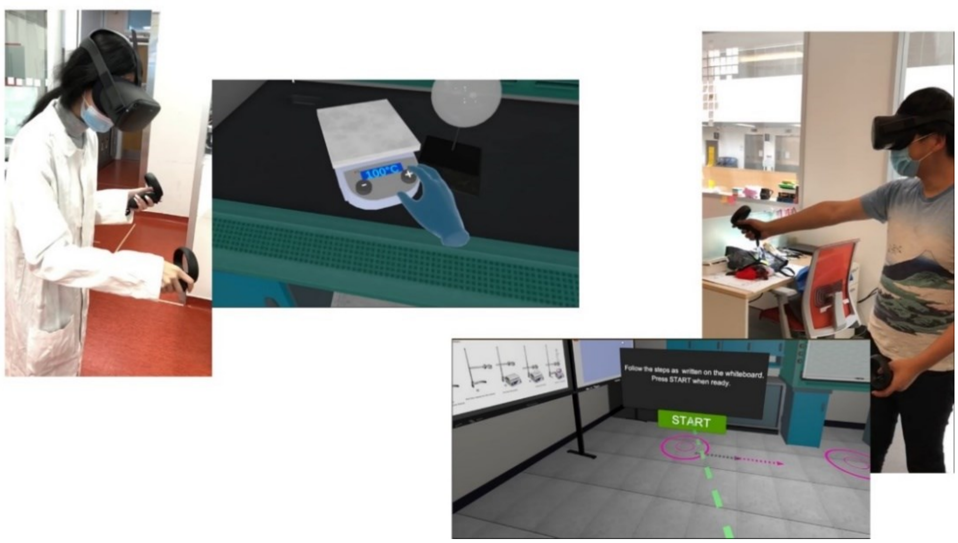


Figure 15.User testing for the pilot study.

Participants interact with equipment in virtual equipment. The green dash-line represents a raycaster (a laser beam emitting from a hand-held controller)

[14] The Workplace Hazardous Materials Information System (WHMIS) is Canada's national hazard communication standard.
https://www.canada.ca/en/health-canada/services/environmental-workplace-health/occupational-health-safety/workplace-hazardous-materials-information-system.html
https://ehs.utoronto.ca/our-services/chemical-and-lab-safety/whmis/whmis-lab-safety-training/

3.2.2. Design

- **Goals:** This pilot study was a fundamental phase of our research process. The purpose of conducting this pilot project was to examine the feasibility of the approach that was intended to be used on a larger scale during the next phase. We used the pilot study to evaluate the feasibility, requirements, evaluation procedures, and implementation of virtual reality completing lab-based tasks. During this phase, we identified needed modifications for larger testing that followed in the next phase.

- **Participants:** Students in Chemistry or any other Science or Engineering program at the undergraduate level were invited to participate. Internet accessibility and willingness to wear a VR headset were the criteria for participation. If they did not own a VR headset, one would be provided to them by mail delivery. Three male and three female university Chemistry students participated in this pilot study.

- **Material:** Our prototype was built for a Chemistry lab using Unity 3D, a popular game engine to build 2D, 3D, and VR games and experiences. It can be modified and exported for different platforms including desktop, mobile, and HMDs. The VR experiment for our pilot study and study 2a was built to be run on HMD and desktop platforms.

    - HMD: Due to the practicality and feasibility of the Android-based mobile headset, the Oculus Quest device (by Meta) was selected as the device of choice for the immersive VR experiment (Figure 16).

Figure 16.Oculus Quest HMD by Meta (former Facebook)

  - Desktop: A Windows desktop version of the VR app was built for the desktop and could run on any laptop or desktop computer with Windows 10 operating system.
  - To minimize the production time a variety of ready-made assets were acquired from Unity and 3rd party asset stores. Other 3D objects were modeled using Autodesk 3Ds Max software.

- **Knowledge Test and Surveys:** A pre-post experiment knowledge test (appendix I, A, and B) was designed using standard lab safety standards. We also collected user interaction data using a built-in data collection system developed by the researchers for this purpose. This was an important component of the system that would help researchers not to rely on third-party application programming interfaces (APIs) due to their limitations or costs. A usability questionnaire (Appendix I, C) was also administered after the completion of the VR experiment by the participants. They were asked to rate their experiment on a 7-point Liker scale (1 very negative and 7 very positive), related to each of the following criteria: Ease of use, memorability, learnability, pleasantness, clarity, visualization, and overall satisfaction, followed by one open-ended questions to ask how they felt about the experience. A sense of presence survey (SUS) [117] (Appendix I, D) was conducted to evaluate participants' sense of presence in IVR. And finally, participants were asked to answer an open-ended question about their experience.

- **Procedure:** Participants would start by completing a demographic survey and a pre-knowledge test (Appendix I, A). Then they were asked to complete a training session in VR to get familiar with controllers and interactions. They would then go to the main area a series of tasks detailed in the next chapter (Table 4) that included locating, grasping, manipulating objects (Figure 17), and ducking. They would also use traveling and teleporting (locomotion) to move around virtual space. (See figures 19-29 in the next chapter that shows the complete procedure). Participants would encounter two accident scenarios in the lab (a small fire and a chemical spill) and would need to react and follow the safety procedures learned in the training area.

The virtual environment for IVR and DVR is the same only the interaction techniques /device type to interact with them differ (VR controllers vs keyboard and mouse).

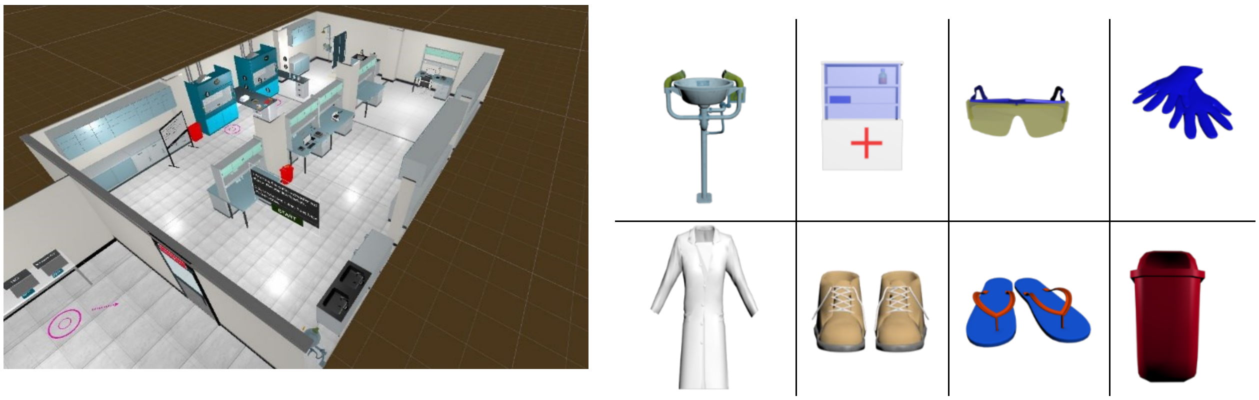


Figure 17.Virtual environment (left) and interactable objects (right)

3.2.3. Results

For the pilot study, only the IVR version was tested on a limited number of university chemistry students (3 males, and 3 females) since the DVR version was not ready at the time. We tested the practicality of using mobile HMD for future studies. This pilot project revealed that participants felt comfortable enough using headsets to complete the experiment. The possibility of motion sickness is a known (small) risk item in IVR.

Despite our concerns about using a joystick for continuous movement and potential cybersickness, no users showed or expressed any symptoms of cybersickness confirming that the app runs smoothly and without the latency that could induce cybersickness. This can be contributed to the measures we took to minimize the potential effect of cybersickness in the following ways:

- Not having any fast-moving objects in the scene
- Having participants in the HMD for no more than 15-20 minutes
- Keeping a high framerate (the maximum possible by an Oculus Quest headset - 72hz),
- Using a “Teleportation” system with a “Fade in-out” effect helps to reduce issues related to motion sickness.
- Requiring no fast movements by the player, also no moving in the virtual environment without the participant's action. We also limit the speed by which users can travel.

A few issues related to usability were identified and rectified. Disabling the ray-cast while hovering on non-interactable objects was added to avoid confusion for users. We also noticed that the recorded data from the headset was not being delivered to the researcher's email which was solved by making sure each headset was connected to a local wi-fi ahead of time. Another performance issue to record videos inside the HMD was identified. This issue was beyond our control as any attempt to record the activity in a VR headset would create considerable latency. As a result, we could not record any live video in the headset while participants were completing the tasks. Since this study and the consequent studies were designed to be conducted remotely due to COVID restrictions, all online surveys and delivery systems were also tested during this study to make sure all components for collecting data functioned as expected.

In terms of the effect of the pilot study on the consequent phases, we can emphasize that it provided us with the knowledge and skills needed to develop the VR experiments for our next studies. It also provided a structure, code base, and proper workflow using VR technology and the Unity game engine for the next phase. In this design-based research, with every iteration of the experiment, we learned to improve the process culminating in a better user experience in the consequent studies. To avoid repetition, since this pilot study formed the core of the large-scale study we provide the details of the experiment, procedure, and evaluation in the next chapter. The opportunity for multidisciplinary collaboration was a major motivator to proceed and extend this pilot study to a full-scale study.

3.2.4. Further Studies

With the completion of the pilot study, we were motivated to investigate the potential values of using VR in STEM education and gained several insights into the requirements, development process, and limitations of conducting the main study with a larger sample size. While improving on the experiment's technical aspects such as eliminating highly reflective materials that would cause flickers, we optimized the textures and lighting effect to improve the VR performance. We also refined the process of downloading and installation of the VR application file on the headset (for first-time VR users). For several users, a pre-installed headset was provided to avoid installation issues. All these measures helped us to prepare for the next study detailed in the next chapter.

Based on the literature review and our pilot study, we answered our research question 1 and established six more research questions (2-7 below) and their related studies:

Literature Review and Pilot Study1: Focused on foundation and motivation

1-What are the most relevant educational theories applicable to a VR framework, and how can we apply them in the design of a VR system in STEM education? This question was answered through a comprehensive literature review detailed in Chapter 2. Although different theories are mentioned in the literature, there was not a clear and applicable linkage between these theories and VR applications. Our literature review helped establish such a link.

Study 2a: Primary focus on interaction and usability

2-Are VR systems more effective in teaching science lab-based procedures compared to the current teaching method (2D / Video)? If so, how immersive and desktop VR compares in interactions, usability, and efficiency in completing assigned tasks?

Study 2b: Primary focus on learning

3-Can using VR and authentic visualization (using ball-stick models) improve learning compared to two-dimensional (2D/Text-based) content?

4- Can flexible interaction type impact efficiency, accuracy, and learning outcome, if it enhances or hinders user experience (UX)?

Study 3: Primary focus on accessibility

5-Can software controls for vertical movement provide improved accessibility for wheelchair users with task efficiency and completion rate similar to non-wheelchair users?

6-Can the improved accessibility (if any) improve learning?

Framework Design: Primary focus on integration and framework

7-What are the specific guidelines for accessibility and general guidelines for designing VR experiences for STEM teaching and learning?

During the three user studies (described in Chapters 4-6), we had 104 participants. The results from 8 participants were discarded due to incomplete experiments or cybersickness.

During the same period, we had the opportunity to (unofficially) use the VR experiment amongst close friends and relatives to gain more insights into the design and development process. However, only the results from the remaining official 96 participants formed the core of the data analysis for this study.

Some of the insights gained during the unofficial user testing were related to the 3D environment i.e., a futuristic game-like environment versus an authentic and realistic-lab VR space. Users' input indicated that while a futuristic environment can be fun and cool like a computer game, it does not reflect an educational environment they are familiar with in the real life. This insight was applied in the development of the base environment and subsequent studies.

Some users suggested improving the lighting of the virtual environment and the contrast between the text and background for better legibility. These suggestions were implemented in studies 2b and 3 using baked lighting and adding more light sources. In the next versions, a customizable text feature will be implemented.

For study 3, wheelchair users suggested adding the height change feature to the handheld controller for convenience. Currently, this feature is tied to a software menu that can be invoked from within the app. A quick test was conducted to add this feature to the joystick where users could change the elevation by moving the joystick up and down. This was the preferred method (compared to software control) by a limited number of users (both wheelchair and non-wheelchair users). However, the result was not included in the document as it was not within the scope of the study.

Chapter 4. Study 2a – Interaction and Usability

In this chapter, we report on the process and findings of Study 2a, focused on general experience, interactions, and usability. For this and all other studies, we follow the standard APA format for reporting user experiments, as described in https://www.scribbr.com/apa-style/methods-section. This chapter starts with an overview of the problem and required features of a science lab for safety considerations. We then describe research questions and hypotheses, the specifications of the VR experiment built for this study, the procedure of user testing present the results followed by a discussion section.

The results of this study were published in IEEE AIVR. 2021 titled: "*ScienceVR*: A virtual reality framework for STEM education, simulation, and assessment."

4.1. Overview

As detailed in the literature review of this document, Learning and teaching Science, Technology, Engineering, and Mathematics (STEM) has unique challenges, in part, due to the abstract nature of concepts in these disciplines and because of the range of issues such as cognitive load, required spatial thinking, and experiential and hands-on learning requirements[28] [35]. Several scientific topics in STEM need a dedicated physical laboratory. These labs often contain toxic and dangerous materials which can present a high level of danger if not maintained and handled carefully. Maintaining such physical facilities and the safety considerations around any science laboratory can be very costly. Research studies suggest that despite the existence of safety standards and guidelines, accidents do happen in higher science education laboratories [39]. While most reported accidents were "relatively" minor, some have caused permanent injuries or fatalities [40]. Lack of knowledge of the established safety procedures and insufficient training were mentioned as contributing factors to systematic safety failures.

The objective of phase two of our study is to determine whether immersive and desktop VR simulations can be used as effective learning tools in science labs, with a focus on learning the safety procedures against the risk of accidental fire and chemical spills. We are using a chemistry lab as an example. In its current format, students are given a standard safety manual for each lab experiment. The instructor explains the safety process before entering the lab in a 10–15-minute class session. Students are then asked to sign a waiver confirming that they read and understood the safety procedure.

We simulated this procedure for three environments, and participants were divided and assigned to each environment separately:

- Desktop VR (DVR)
- Immersive VR (IVR), using a head-mounted display (HMD)
- 2D (Text/Video-based- A video link is provided on page 69 of this document)

As mentioned in section 1.3, our specific questions for this study were:

RQ2: Are VR systems more effective in teaching science lab-based procedures compared to the current teaching method (2D / Video)? If so, how immersive and desktop VR compares in interactions, usability, and efficiency in completing assigned tasks?

We broke down this question into the following, more detailed, ones:

- 2a) Are VR systems more effective in teaching science lab-based procedures compared to the current teaching method (2D / Video)?
- 2b) If so, how immersive and desktop VR compares in interactions, usability, and efficiency in completing assigned tasks?
- 2c)-How to implement an in-app data collection system without relying on external/3rd
- party tools to measure user performance? What are the advantages of such a system?
- 2d)-What is the impact of immersion and sense of presence on the engagement level and learning outcomes?

To answer the above question, we developed a virtual lab with a safety procedure activity and two accidents scenario. Figure 16 shows the virtual environment and some of the interactable objects that users will see and interact with.

Users were tasked to complete several common lab activities compiled by the domain expert (our collaborators). The measurements are based on the primary goal (learning) plus standard metrics such as efficiency, accuracy, and general usability. Some of these criteria are measured through objective data and some are subjective, as shown in Table 2. The subjective usability test included Ease of use, Memorability, Learnability, Pleasantness, Clarity, Visualization, and Overall satisfaction. We also included a set of presence questions for the immersive VR.

Table 2. Research Questions and Variables

Dependent Variable	RQ #	Type	Instrument
Learning outcome (Safety procedure)	2a	Objective	Knowledge test Observation
Usability	2b	Subjective	Survey
Task completion (Grasp, Find, verify, travel, apply)	2c	Objective	In-app data
Efficiency and accuracy	2a & 2b	Objective	In-app data
Presence (for immersive VR)	2d	Subjective	Survey

4.2. Participants

Students in Chemistry or any other Science or Engineering program at the undergraduate level were invited to participate. Internet accessibility and willingness to wear a VR headset were the criteria for participation.

Participants were divided into three groups to complete the VR experience before the real lab[15]:

- Group A (2D), 12 participants
- Group B (DVR), 12 participants
- Group C (IVR), 12 participants

Participants needed to have access to a computer with the Internet and were required to download the VR program. IVR participants were informed about the potential small risk of fatigue and motion sickness. They were provided with the equipment if needed, following the health and safety regulations (contactless drop-off and clean-up).

Due to the remote nature of the study, a detailed instruction manual was provided to participants that described what the study entailed, including instructions on how to stop the study at any time. Participants were asked to read the informed consent document and to sign if they agree to participate.

Ethics approval was received from the Carleton University Ethics committee. Participants received a $10 gift card as a token of appreciation for their participation in the study. 38 volunteers participated in this study, 22 were males and 16 were females.

[15] The ideal condition would be to offer this VR experiment to students once before the real lab and study the same process in a real lab and compare the results. However due to the restriction imposed under pandemic the real lab experiment was not done.

4.3. Material

For this study, the prototype environment was the enhanced version of the base environment detailed in Chapter3. All hardware and software used in this study were the same as detailed in Section 3.2.2. The only difference with the base environment used for the pilot study was the added DVR version for desktop users in this study. We also fixed all the technical issues such as missed interactable objects and typing errors in this version.

From the production and development point of view, we kept the workflow simple to allow flexibility yet to maintain the mood and feel of a science lab as realistic as possible. Figure 18 shows the image of an actual lab (left) vs our virtual environments (right).

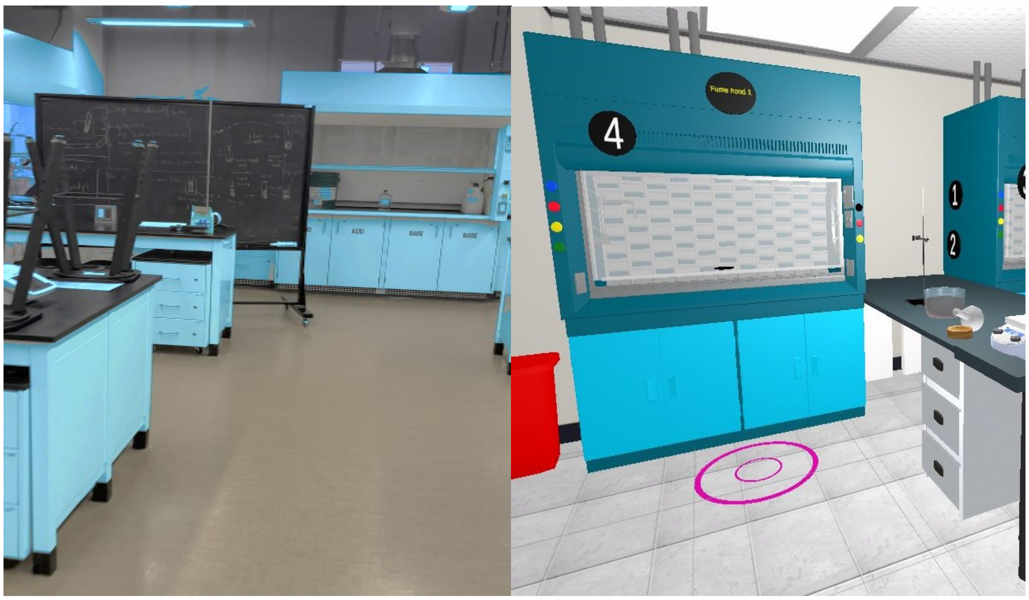


Figure 18.Physical lab (left) vs. virtual lab (right)

Our proposed environment includes two separate areas described in more detail in the next section. Since our target audience could have limited or no experience using IVR, we have included a tutorial activity to help them gain experience using touch controllers. This was an important design consideration based on scaffolding learning theory to help users acclimatize themselves and avoid potential motion sickness for some users.

Participants in all three groups were asked to complete a survey asking demographic questions including age, gender, handedness, and familiarity with VR. In addition to this survey and the required VR software, the following data collection instruments were used:

- **Knowledge tests:** Pre and Post Knowledge Tests (see appendix I) were designed using questions from the WHIMIS safety manual prepared by the collaborator of this study, a chemistry instructor. The pre-experiment questions were more general and covered theoretical aspects while the post-questions were more practical and related to specific items in the training material. They can be found in Appendix II and III.
- **An in-app data collection system:** A built-in data collection system of IVR and DVR was used to track and record users' interactions.
- **Usability survey:** A post-study survey was conducted to evaluate participants' experience and sense of presence. They were asked to rate their experiment on a 7-point Liker scale (1 very negative and 7 very positive), related to each of the following criteria: Ease of use, memorability, learnability, pleasantness, clarity, visualization, and overall satisfaction, followed by one open-ended questions to ask how they felt about the experience.
- **Sense of presence survey:** A Slater Usoh Steed (SUS) questionnaire [117]. SUS was selected since it is more consistent compared to other methods. The SUS questionnaire contains the following questions, and it was administered to evaluate participants' sense of presence in IVR:

1. Did you have a sense of "being there in the virtual environment"?
2. Where there were times during the experience when the virtual environment was a reality for you?
3. When you think back to the experience, do you think of the virtual environment more as images that
4. you saw or more as somewhere that you visited?
5. During the time of the experience, which was the strongest overall, your sense of being in the virtual
6. environment or of being elsewhere?
7. Do you think a virtual environment is a place in a way similar to other places you've been to today?
8. During the time of your experience, did you often think to yourself that you were actually in the virtual environment?

In this study and all subsequent ones, Survey monkey (surveymonkey.com) was used to create online surveys and make them available to participants for online data collection. We also offered an open-ended question on how participants felt about the experiment to gain more insight into their experience. The learning evaluation was based on the difference between pre-and post-knowledge scores (delta). The task completion and efficiency were also considered but since 2D users had less VR exposure, we admit that the result provides limited insight. The usability evaluation is based on a combination of usability criteria (ease of use, learnability, memorability, pleasantness, visualization, and overall satisfaction) and subjective satisfaction (likability) introduced by Nielsen [66]; as well as the ISO 9241-11 standard, which interprets usability as "Effectiveness, Efficiency, and Satisfaction. Furthermore, the PACMAD model has additional factors (to the ISO model) that include learnability, memorability, errors, and cognitive load [67].

By collecting objective data, we could validate if users were completing the assigned tasks, and what would the success and error rates be. For objective data (effectiveness and efficiency), we used the logged data for the interactions to record the data automatically during the experiment. This feature will eliminate errors in data collection and will not interrupt the process as participants will not experience any intermittent distraction. More detail on data collection systems is provided in the next section.

Based on seven VR design elements [53] (basic interaction, immediate feedback, virtual rewards, realistic surrounding, use of audio feedback or narration, moving around, and knowledge test) and related learning theories described in Chapter 1, the following design features were implemented in this experiment:

- Help buttons containing information and hints appear when hovered by a pointer
- Highlighted knobs and gauges with visual outlines
- Traveling and teleporting features so that users can move around the space within the safe area defined in the headset
- Haptic feedback for every intractable object (within hand reach) using handheld controllers
- Use of the ray-cast for remote objects mainly for 2D UI's/help elements
- Use of snaping and drop zone feature to guide user's action
- Use of virtual rewards
- Use of in-app and external knowledge test

4.4. Procedure

The experiment for each group had two parts:

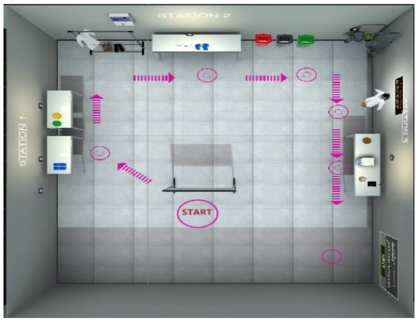
a) training part using the selected environment. This was the main part of the research and included pre- and post-knowledge tests to evaluate learning and knowledge improvement, in addition to a post-experience usability survey.

b) practical test using VR. This part should have been done within a real laboratory however, due to COVID restrictions, we tested IVR and DVR groups using their own environment. The 2D group was randomly divided into two sub-groups to be tested by IVR and DVR after a short training on VR systems. This part was used to evaluate task effectiveness and efficiency using the in-app data, but we understand the limitations due to simulation. A presence survey was conducted at the end for IVR users.

In VR, participants spent 20-30 minutes completing the required tasks. Users' journey in each step is described through the image sequence in Figure 19 (a-k). Each task shown in Table 3 included one or two lower-level subtasks. For example, task 1 of grasping an object is composed of three subtasks in IVR: looking around, choosing an object (by pointing a ray cast, or extending the arm, and "grasping" gestures. In DVR on the other hand the same task can be achieved by clicking on the object to add it to the inventory confirming that the object is "grasped".

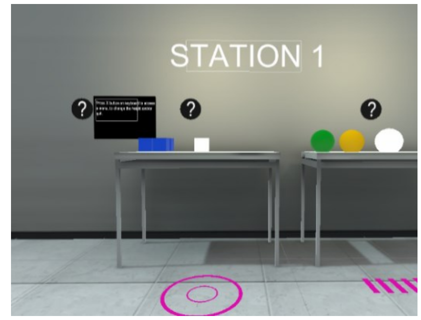
Table 3. Complete List of Tasks

	Task	Objects to interact with or identify
1	Grasp (Subtask: Select)	1. Lab coat 2. Safety goggle 3. Glove
2	Verify (Quiz) (Subtask: Check)	1. Shoe 2. Pants 3. Hair 4. Contact lens
3	Locate (Subtask: Identity)	1. Sink 2. Eye-washer 3. Shower 4. Red bin 5. Green bin 6. Black bin
4	Travel to (In IVR) Click On (In DVR)	1. Sink 2. Eye-washer 3. Shower 4. Red bin 5. Green bin 6. Black bin
5	Manipulate	1. Press the "Start" button to run an interactive science experiment and respond to the events such as chemical spills and a small fire

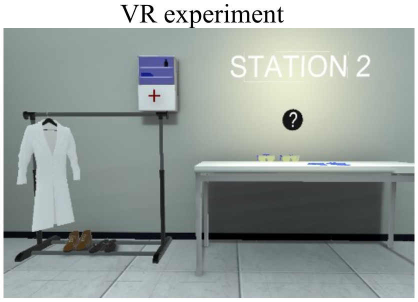


(a) Top view, Tutorial area, user wears the VR headset and starts the VR experiment

(b) User is assigned an automatic ID and is guided to the next steps by voice and text



(c) User learns simple interactions using controllers by stacking cubes and moving spheres

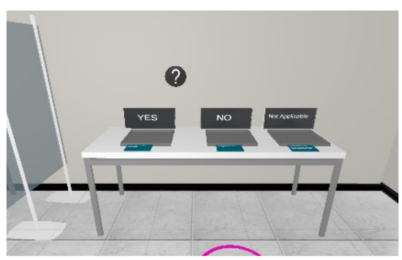


(d) User learns about safety gear, lab coat, gloves, safety glasses, first aid kit

(e) User learns about disposing of chemical waste and use of shower and eye-washer

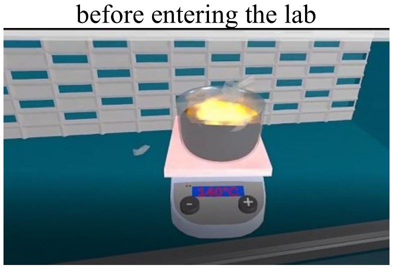
(f) User learns about personal effects (contact lens and jewelry) protocols and how to interact with virtual equipment. The tutorial section is completed.

(g) User enters the next room and is tested on using proper safety gears before entering the lab



(h) User will answer a few questions about the use of contact lenses and jewelry

(i) User enters the lab for the main experiment



(j) User reacts to two accident scenarios (small fire and chemical spill)

(k) User walks to the shower/eye washer and applies water. The VR experiment is complete. The user exits the VR lab.

Figure 19.User journey through VR space

Participants started by completing a demographic survey and pre-knowledge test. Participants were then asked to go through one round of training and then complete one round of practical tests within DVR/IVR app as shown in the previous section.

As an immersive VR experience can overwhelm first-time users, participants were guided to interact with simple objects, travel, or teleport (locomotion), pick up objects, and use help tips. This level was built based on the scaffolding principle in learning theories to help participants build the skills and knowledge necessary to navigate and interact with objects. It also helped to avoid cognitive load. Help tips were hidden under a question mark icon and were revealed when users point at them with the controller using the ray cast. This helped to create a less cluttered environment.

The next area, divided into two rooms contained more interactable objects including personal protective gear, safety questions, and scientific experiment stations known as fume hoods. Participants were guided through signage and guiding tips to complete tasks that included finding/locating objects, touching/picking up objects, and reacting to accident scenarios.

After completing the training in areas 1 and 2, the VR participants were asked to complete the test tasks in the second area without showing them any of the help tips or guides to complete the experiment. They were informed that the accuracy and efficiency (speed) of their performance would be recorded.

The 2D group was asked to read a written lab manual and watch a video with training content (Figure 20). This group was later divided randomly and assigned to either DVR or IVR environment to conduct one basic VR training and the in-app practical test.

All participants completed the post-knowledge test and usability survey for their training environment at the end.

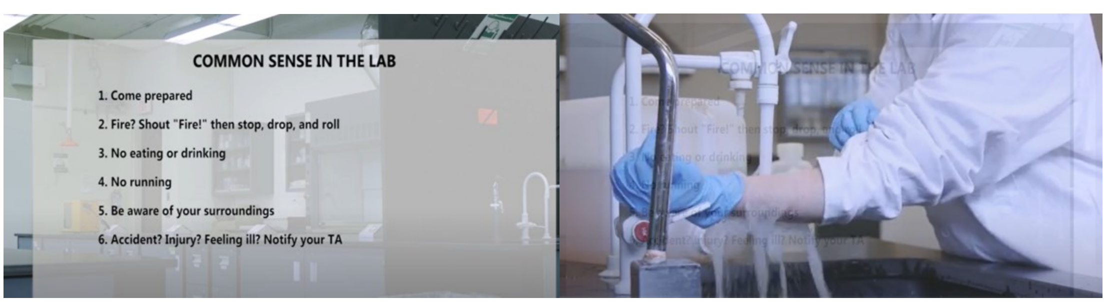


Figure 20.Screenshots from the safety procedure video[16]

[16] By University of Toronto describing the lab safety protocols (Ref: https://www.youtube.com/watch?v=BGSVluvZMgY&t=201s)

For this study we have developed the following hypotheses:

- **Hypothesis 1:** We hypothesize that the learning outcome will be positively impacted by the training environment. The IVR and DVR groups who were trained in each respected area will score higher in the post-knowledge test indicating a better learning outcome compared to 2D. We expect the IVR score to be the highest.
- **Hypothesis 2:** We expect that for the usability across three platforms, the 2D condition (in the form of 2D/ video) will rank higher than IVR and DVR for ease of use, and the IVR will be in second place. However, for other areas such as learnability, visualization, memorability, and overall satisfaction IVR and DVR will rank higher.
- **Hypothesis 3:** For the combination of efficiency and accuracy in task completion, we hypothesize that DVR will rank higher than IVR due to the familiarity of users with the interactions and desktop environment. We cannot test for efficiency and accuracy in 2D as we have no access to physical labs.
- **Hypothesis 4:** We hypothesize that IVR will create a strong sense of presence and immersion, creating an engaging experience where participants stay longer in the simulation. It will positively impact learning.

4.5. Results

4.5.1. Demographic Information

Originally 38 volunteers participated in this study, 22 were males and 16 were females. Their average age was 25 with a standard deviation of 6.45. Two participants were excluded from the results due to an incomplete experiment caused by motion sickness in VR and an incomplete post-experiment test and survey.

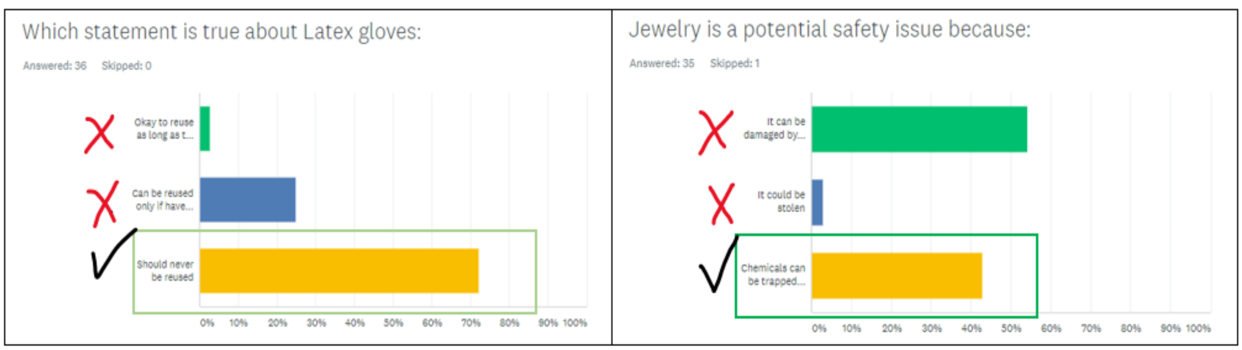
86% of the participants were right-handed and 14% were left-handed. All of them were university students in Chemistry or other science/engineering program. On average they had taken 5.10 chemistry courses with a standard deviation of 3.70.

94% had already completed WHIMIS safety training. 66% had prior experience in immersive VR with a variety of games including “Beat Saber”, “Super-Hot demo”, 360 degrees experiences e.g. roller coaster, flying in a plane, “Walking dead” etc. The remaining 34% were aware of VR but had never experienced it.

4.5.2. Learning Outcomes

The learning outcome for three environments was analyzed based on pre-post knowledge test results. Scoring was done based on a point system for a pre-post knowledge test out of 9 points with one point given to each correct multiple-choice, short-answer question, and fill-in-the-blank question; half a point for partially correct answers on short-answer questions and zero points on skipped questions or incorrect answers. The short answer questions were scored according to a rubric that indicated the words and phrases required for 1 point or half a point.

Although the majority of participants had already completed safety standard training, the pre-knowledge test on lab safety revealed that there are noticeable gaps in their understanding and recollection of the knowledge judging by the number of incorrect answers by the participants. For example, in Q3: "To reduce the chance of accidents", 36% of the participant gave the wrong answer of "All of the above", while the correct answers were "Use personal protective equipment (PPE)" and "Use the smallest quantity of material necessary", 23% of participants were not sure what items would NOT include in personal protective equipment and 14% answered incorrectly to the protocol of re-using gloves. 58% did not know the reason Jewelry is a potential safety issue in a lab (Figure 21).



The correct answer is: "Should never be used" The correct answer: "Chemicals can be trapped under, react"

Figure 21.Answers to Q3 and Q4

(X denotes incorrect answers to 2 questions in the pre-knowledge test)

On the question of "Safety regulations about the use of contact lenses" 25% of participants gave the incorrect answer (Figure 22):

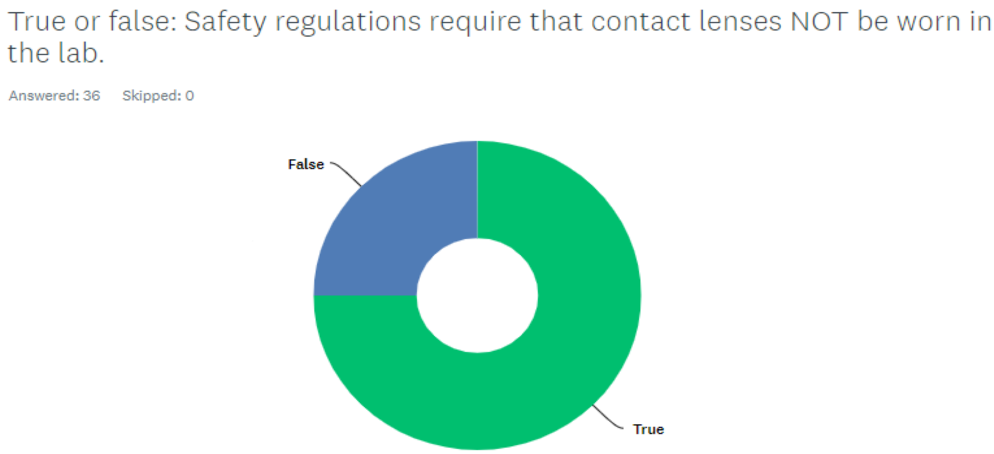


Figure 22.Safety regulation answers on contact lenses

And on the question of “What to do first in case of an accident in the lab?”, only 55% answered correctly “informing/calling the instructor. Participants’ answers revealed that there is a disconnect in their knowledge of safety protocols. The current safety training is based on a text-based safety manual that students are obligated to read and remember.

We performed three sets of one-way independent analyses of variances (ANOVA) with test scores across all three groups. ANOVA on the Pre-Knowledge Test showed no significant difference, indicating that all groups were at the same level of prior theoretical knowledge. But for Post and Delta (Post minus Pre), ANOVA showed significant differences, and further Tukey HSD post hoc test showed significant differences in the cases of IVR vs. 2D and DVR vs. 2D, with no significant difference for IVR vs. DVR. Table 6 shows the mean and standard deviation for all sets of score data and Figure 33 shows the ANOVA results. While the mean values are similar for the pre-test, VR participants scored higher in the post.

This result supports our first hypothesis (H1) that the IVR and DVR training has a positive effect on learning outcomes as indicated by the higher score in the post-knowledge test. However, there is no significant difference between IVR and DVR in this regard. This result does not support our expectation that IVR will perform better than DVR. In fact, the mean post score for DVR is slightly higher (although not significant).

ANOVA was conducted on the post-test scores of the three groups with gender types as covariates (Table 4). With the value of $F(2,33) = 1.436$, $p < 0.240$, the effect of gender on the difference between the pre and post-knowledge tests was not significant.

Table 4. Knowledge Test Mean Score and Standard Deviation for 2D, IVR, and DVR

	2D	IVR	DVR
Pre	M = 6.06 SD = 1.48	M = 6.71 SD = 0.98	M = 7.15 SD = 1.06
Post	M = 3.00 SD = 1.26	M = 7.19 SD = 1.98	M = 7.46 SD = 2.53
Delta	M = -3.05 SD = 1.91	M = 0.48 SD = 2.56	M = 0.31 SD = 2.34

ANOVA Summary Independent Samples k=3

Source	SS	df	MS	F	P
Treatment [between groups]	7.3261	2	3.663	2.57	0.091770
Error	46.9968	33	1.4241		

Tukey HSD Test

This test will be performed only if K>2
and the analysis of variance yields a
significant F-ratio.

M1 = mean of Sample 1
M2 = mean of Sample 2
and so forth.

(a)

ANOVA Summary Independent Samples k=3

Source	SS	df	MS	F	P
Treatment [between groups]	149.941	2	74.9705	18.85	<.0001
Error	131.2471	33	3.9772		

Tukey HSD Test

HSD[.05]=2; HSD[.01]=2.55
M1 vs M2 P<.01
M1 vs M3 P<.01
M2 vs M3 nonsignificant

M1 = mean of Sample 1
M2 = mean of Sample 2
and so forth.

(b)

ANOVA Summary Independent Samples k=3

Source	SS	df	MS	F	P
Treatment [between groups]	95.064939	2	47.532469	9.11	0.000708
Error	172.252825	33	5.219783		

Tukey HSD Test

HSD[.05]=2.29; HSD[.01]=2.92
M1 vs M2 P<.01
M1 vs M3 P<.01
M2 vs M3 nonsignificant

M1 = mean of Sample 1
M2 = mean of Sample 2
and so forth.

(c)

Figure 23.ANOVA Results for learning outcome (a) pre, (b) post, (c) delta

Samples 1, 2, and 3 correspond to 2D, IVR, and DVR.

4.5.3. Usability

A "Likert" scale survey (1=lowest 7=highest) on the ease of use, memorability, learnability, pleasantness, clarity, visualization, and overall satisfaction for each environment was administered. Table 5 shows the mean and standard deviation for all usability data across three groups. Based on the mean values, participants found 2D easier to use which is expected, but in other questions, IVR/DVR seemed preferred.

Table 5. Usability Mean Score and Standard Deviation for 2D, IVR, and DVR

	2D	**IVR**	**DVR**
Q1 (Ease of use)	M = 6 SD = 0.6	M = 4.8 SD = 1.1	M = 4.3 SD = 1.7
Q2 (Memorability)	M = 4.7 SD = 0.7	M = 5.8 SD = 1.5	M = 5.7 SD = 1
Q3 (Learnability)	M = 4.5 SD = 0.7	M = 5.6 SD = 1.6	M = 5.6 SD = 0.9
Q4 (Pleasantness)	M = 4.4 SD = 0.8	M = 4.7 SD = 1.4	M = 5.3 SD = 1.2
Q5 (Clarity)	M = 4.8 SD = 0.6	M = 5.2 SD = 1.6	M = 5 SD = 1.3
Q6 (Visualization)	M = 4.3 SD = 0.5	M = 5.6 SD = 1.3	M = 5.9 SD = 0.9
Q7 (Overall Satisfaction)	M = 4.8 SD = 0.6	M = 4.8 SD = 1.1	M = 5.3 SD = 1.2

Figure 24 shows side by side comparison of the results for ease of use, memorability, and learnability in each group. The result confirms our hypostasis 2 (H2) that the 2D condition ranked higher in ease of use than IVR and DVR but scored lower in the area of memorability and learnability.

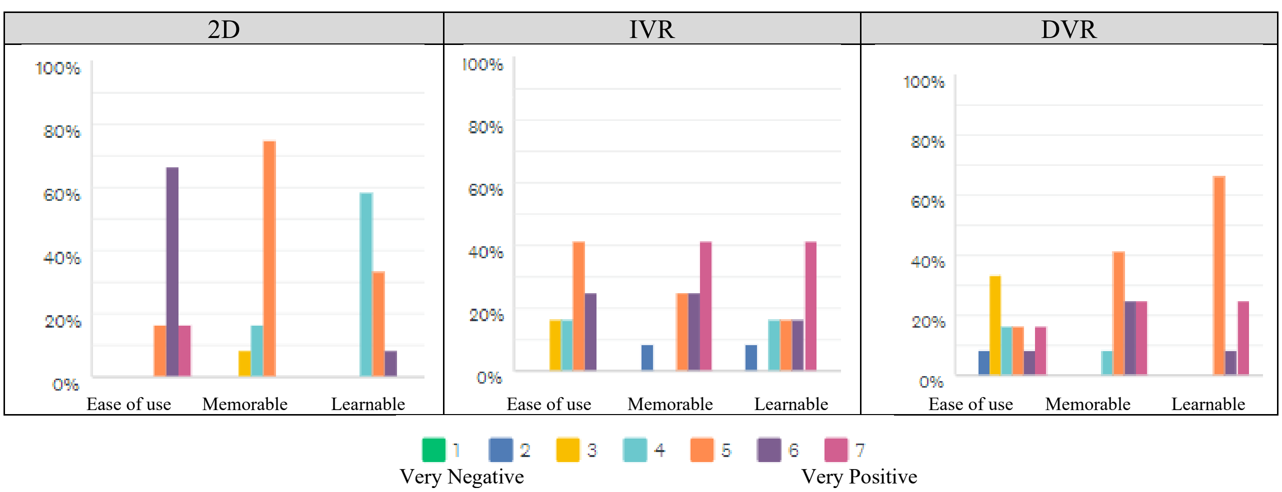


Figure 24.The ease of use, memorability, and learnability percentage for the three groups

Figure 25 shows side by side comparison of the results for pleasantness, clarity, and visualization in each group. The result confirms our hypostasis 2 (H2) for the three areas in IVR and DVR.

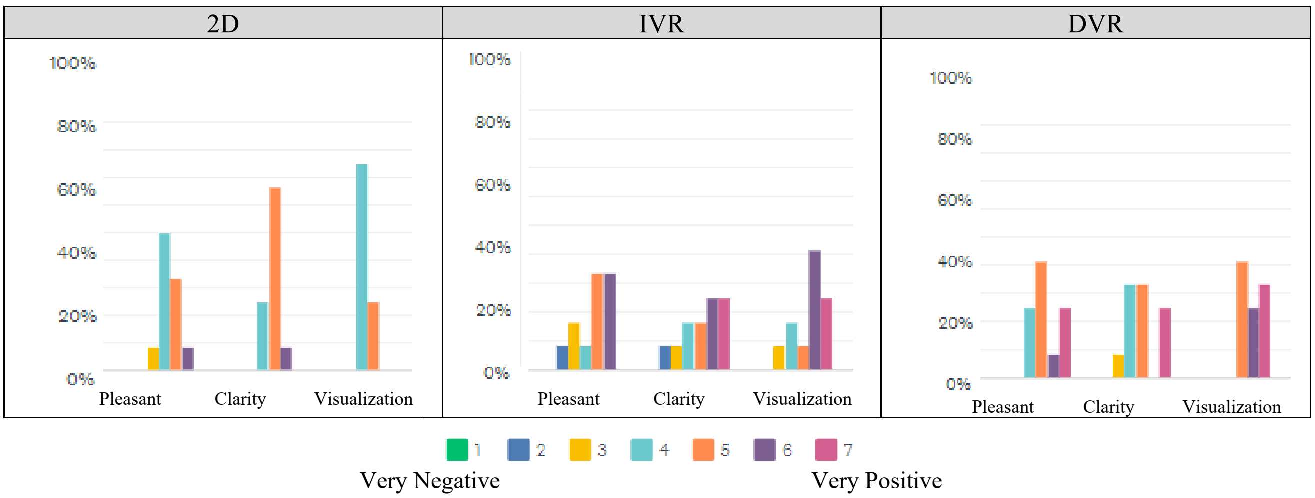


Figure 25.The pleasantness, clarity, and visualization percentage for the three groups

And finally, on the overall satisfaction rate between the three groups, Figure 36 shows the overall satisfaction percentage indicating lower overall satisfaction in the 2D group compared to the other two groups confirming our H2.

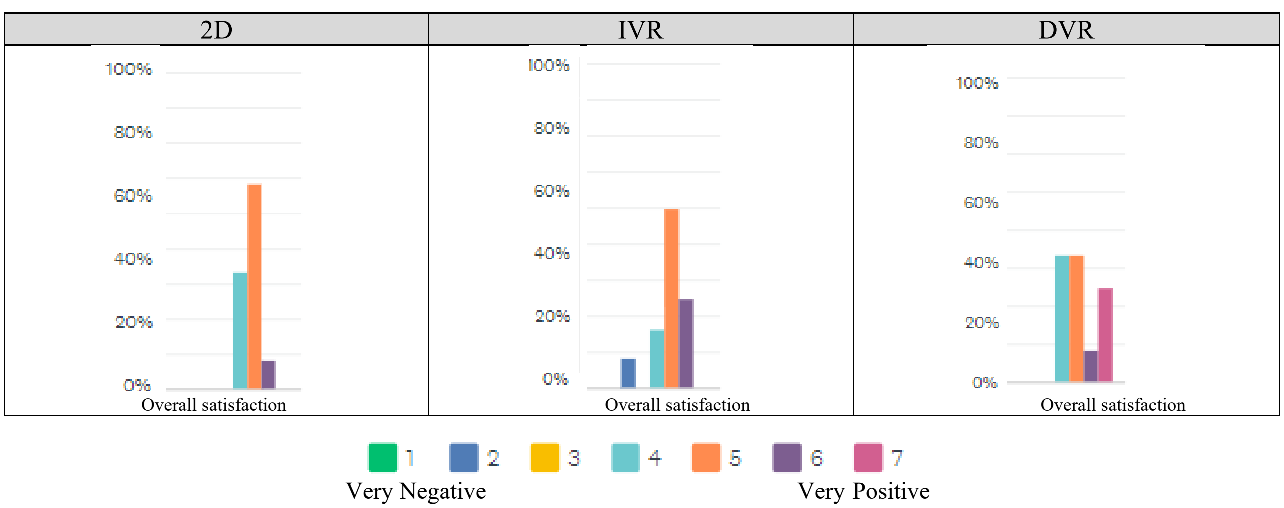


Figure 26.Overall satisfaction for each group

For usability, a non-parametric test (Kruskal-Wallis, KW), as well as a paired test (Mann-Whitney, MW), was conducted due to the non-parametric nature of our Likert choices and data. KW and MW are non-parametric versions of ANOVA and T-Test. Due to the lack of standard post-hoc tests for KW in the tools we used (http://vassarstats.net/), we chose to apply MW on pairs of samples to investigate pairwise differences[17]. The result shows the differences in seven areas of usability (Table 6). IVR and DVR showed no significant difference in any of the seven questions which were shown in their mean scores as well.

- For Ease of Use, Kruskal-Wallis and paired post hoc test showed significant differences between IVR/DVR and 2D. This was the only case where 2D outperformed VR.
- For Memorability, Kruskal-Wallis showed an overall difference and Mann-Whitney confirmed that IVR was performing better than 2D, although IVR vs. DVR showed no significant difference.
- For Learnability, Kruskal-Wallis showed an overall difference, and the paired test showed a significant difference between IVR/DVR and 2D.
- For Pleasantness and Clarity, there were no significant differences.
- For Visualization, there was a significant difference between IVR/DVR and 2D, as shown by Kruskal-Wallis and Mann- Whitney test. The difference between IVR and DVR was not significant.
- For Overall Satisfaction, Kruskal-Wallis did not show any overall significant difference.

[17] Using t-test after ANOVA is not recommended as it may increase the probability of false positives/negatives. Insteda, standard post-hoc tests are done after ANOVA to investigate pairwise differences. Using MW instead of a post-hoc after KW is found acceptable by our advisor statistics expert and other sources: https://www.ibm.com/support/pages/post-hoc-comparisons-kruskal-wallis-test

Table 6. Kruskal-Wallis and Mann-Whitney Test for 2D, IVR, and DVR

-Level of Significance for the test was 0.05
-Lower limit =42 Upper limit =102

	H	df	P	Mann-Whitney Post Hoc Test
Q1 (Ease of use)	9.51	2	0.0086	M1 vs M2 P=0.0051 U=23 M1 vs M3 P= 0.0164 U=30 M2 vs M3 P= 0.4179 U=57.5
Q2 (Memorability)	8.76	2	0.0125	M1 vs M2 P=0.0078 U=25.5 M1 vs M3 P= 0.0244 U=32.5 M2 vs M3 P= 0.4533 U=58.5
Q3 (Learnability)	7.85	2	0.0197	M1 vs M2 P=0.0404 U=36 M1 vs M3 P= 0.0067 U=24.5 M2 vs M3 P= 0.7718 U=66.5
Q4 (Pleasantness)	3.35	2	0.1873	M1 vs M2 P=0.4009 U=57 M1 vs M3 P= 0.0643 U=39.5 M2 vs M3 P=0.4179 U=57.5
Q5 (Clarity)	0.58	2	0.7483	M1 vs M2 P=0.3843 U=56.5 M1 vs M3 P= 0.9283 U=70 M2 vs M3 P= 0.7039 U=65
Q6 (Visualization)	13.59	2	0.0011	M1 vs M2 P=0.0131 U=28.5 M1 vs M3 P= 0.0002 U=7.5 M2 vs M3 P= 0.749 U=65
Q7 (Overall Satisfaction)	0.35	2	0.8395	M1 vs M2 P=0.5222 U=60.5 M1 vs M3 P= 0.749 U=66 M2 vs M3 P=0.726 U=65.5

(Samples M1, M2, and M3 correspond to 2D, IVR, and DVR respectively)

An open-ended question was included in the post-experiment survey to ask how participants felt about their experience. The answers revealed that IVR groups had a mostly positive experience. Some of the comments were: "I enjoyed this experience. It is useful and the user would enjoy his time while learning. This is the first time I use a VR-based lab and it seems fun and realistic" and "overall it was a positive and very neat experience".

25% of the IVR participants pointed out that they had "experienced slight motion sickness due to teleportation" or commented on the difficulty of "using a dustpan" or moving the "broken glasses", into the bin.

Another participant mentioned being very dizzy and needing to take a break to be able to continue with the experiment. One participant stated that getting "slight dizziness" using the joystick. According to this participant "It did not bother me as I got used to it quickly. I liked the experiment very much".

Similarly, DVR participants had an overall positive experience. Several participants commented on how realistic the environment was. A few participants commented on the difficulties of "picking up objects such as "hotplate", "beaker" or other objects and rotating them. Two participants commented that moving in the DVR environment made them "a bit dizzy".

Another participant stated that "this was a very engaging experience". Several participants commented that the experience was like a game and that it was "really fun" and enjoyable saying "I enjoyed it and think it is a good way to learn safety protocols". Several participants mentioned that it was difficult for them to use the combination of mouse and keyboard to grasp, rotate and move objects.

Only a few participants from the 2D group answered the open-ended question commenting on how it was easy to run and watch the video but difficult to read and concentrate on the text version.

We conclude that the students had a more positive experience and feelings about learning in immersive and desktop VR than learning from a 2D method.

4.5.4. Efficiency /Task completion (In-app Data)

Looking at the data generated by the participants' interactions in IVR and DVR we first made a list of questions to create an overview of how participants performed inside the IVR and DVR and how they interacted with the environment.

These questions included: How many participants in each group conducted and completed the training and testing level (by gender)? How long did it take on average to complete the tasks? Type and frequency of interactions (Hover, Select) and average duration for the main interactions.

In each group, we had 12 participants. (7 M, 5 F). For the duration, four major time factors were measured: Total time/duration spent in each environment (in Minutes), the number of data points generated (via the interaction i.e. hover or select), duration spent in the training area1, duration spent in training area2, and duration spent in the testing area.

A one-way ANOVA was conducted to analyze the difference between IVR and DVR. On average participants spent about 25.50 minutes in IVR simulation compared to 20.36 minutes for DVR. The average number of data points generated in IVR was 340 interactions vs 165 for DVR. A higher number of interaction data can be interpreted as more attempts and exploration to interact with objects.

Although ANOVA does not show a significant difference in the two environments in the duration and the generated data points (F =0.26, p= 0.62), further data analysis revealed that 60% of the IVR participants fully completed both training and testing areas/tasks (68% almost, with minor details missing), while only 33% of DVR participants completed both (50% almost completed).

This justifies the fewer number of generated data points and the lower usability score in the DVR compared to the IVR. The completion percentage among female participants in IVR was slightly higher than males with 57%. In the DVR simulation, this percentage was 75% for male participants and 25% for female participants.

Table 7 shows the average and standard deviation of the four major time factors extracted from in-app data.

Table 7. Duration and Generated Data Points

Units (M) = Minutes

	IVR					DVR				
	Duration in VR (M)	No. of Data points generated	Time spent in training area 1(M)	Time spent in training area 2 (M)	Time spend Testing (M)	Duration in VR (M)	No. of Data points generated	Time spent in training area 1(M)	Time spent in training area 2(M)	Time spend Testing (M)
Mean	25.50	340.08	9.44	11.73	4.19	20.36	165.25	6.55	9.93	3.61
Std Dev	10.90	92.61	5.56	5.17	2.98	11.42	80.97	4.59	5.54	2.56

	IVR from 2D					DVR from 2D				
Mean	12.05	140.33	2.30	0.00	9.34	11.10	98.00	2.58	0.00	8.57
Std Dev	5.66	65.30	2.57	0.00	3.62	3.15	17.04	1.51	0.00	2.48

Since we could not test in a physical lab, we randomly set the 2D group to IVR and DVR to test their simulated lab performance. A similar pattern was recorded among 12 participants (7 M, 5F) who came from a 2D experiment. 6 participants (3 Male and 3 Female) were assigned to IVR which we identify as the “IVRfm2D” group, and 6 other participants (3M, 3F), were assigned to DVR, i.e. “DVRfm2D”. These groups had the basic in-app training to learn how the buttons on the controllers work and were asked to immediately complete the in-app testing phase without any other training. The completion rates for IVRfm2D and DVRfm2D were 83% and 50% (almost completed), respectively. IVRfm2D is slightly higher than IVR which can be due to the extra time they took, the relatively simple nature of tasks that didn’t need significant training, a random effect due to the small sample size, and possibly some missing elements in VR training that were covered in 2D. The results (Table 4.6) show that it took the 2D participants much longer to do the practical test, but we cannot count this as evidence of better training for VR participants, as the longer time could simply be the result of not being familiar with the VR simulation.

From the data review of the IVR and DVR participants, we can see that the IVR group is performing better than the DVR group in terms of conducting the task with more efficiency and accuracy. However further analysis for this part is required and is in progress.

In summary, the result is inconclusive to confirm or contradict our hypothesis 3 (H3) that "for the combination of efficiency in task completion, we expect that DVR will rank higher". As it currently stands IVR is showing a better performance.

4.5.5. SUS presence survey

On the sense of presence in the virtual environment, each question was separately analyzed. It is worth noting that due to the covid-related restriction we did not have an opportunity to conduct the same experiment in the real-world lab to be able to compare the sense of presence in an equal condition to the virtual lab.

For Q1: "Did you have a sense of "being there in the virtual environment?" participants were asked to rate their sense of "being in the virtual environment", on a scale of 1 to 7, where 7 represents the normal experience of being in a place. Result shows that 75% of participants reported a strong sense of "being there" in the immersive virtual lab (Figure 27).

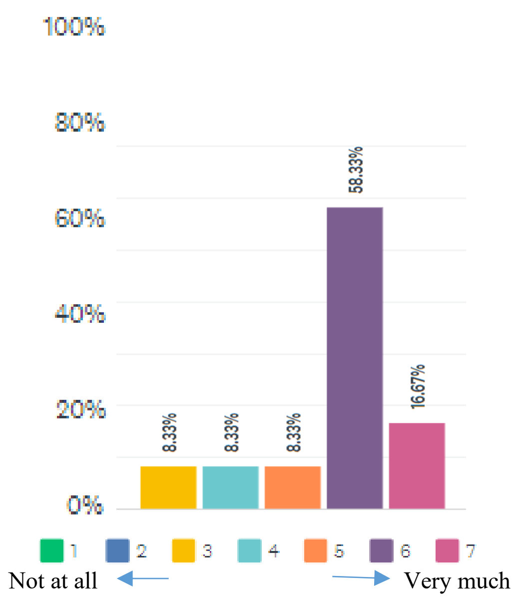


Figure 27.Q1 of the SUS survey for the sense of presence

For Q2: "Where there were times during the experience when the virtual environment was a reality for you?", over 50% of the participants reported that the IVR environment was a reality for them "almost all the time" suggesting that the IVR environment was close to an authentic science lab environment (Figure 28).

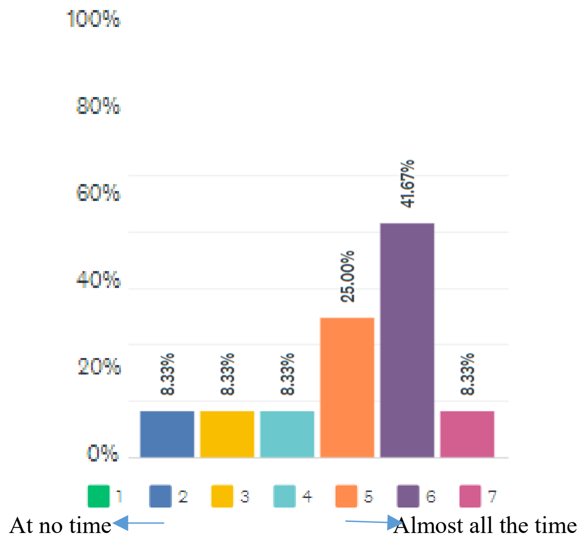


Figure 28.Q2 of the SUS survey for the sense of presence

For Q3: "When you think back to the experience, do you think of the virtual environment more as images that you saw or more as somewhere that you visited?" Over 50% of the participants thought of the virtual space as a place/location they had visited and not just an image they saw indicating the virtual space felt "real" for half of the participants. Several comments by the participants also confirm this notion of how "realistic" it was.

For Q4: During the time of the experience, which was the strongest overall, your sense of being in the virtual environment or of being elsewhere?" 75% percent of participants had the feeling of being elsewhere.

For Q5: Do you think the virtual environment as a place in a way is similar to other places you've been to today?", Over 58% of participants strongly felt the virtual environment was similar to other places they have been.

For Q6: During the time of your experience, did you often think to yourself that you were actually in the virtual environment? 41% of participants often thought of themselves to be in the VR environment.

The overall results indicate that there is a correlation between the feeling of "being" in the virtual environment and the similarity to the "real" world could be an attraction that encouraged participants to stay longer in the virtual environment and complete the tasks that in turn reflected as the higher completion rate and higher post knowledge test.

This would confirm our hypothesis 4 that the IVR condition will create a strong sense of presence and immersion, creates an engaging experience where participants stay longer in the simulation, and positively impact the learning outcome.

4.6. Discussion

Overall, and in response to our research questions, the results from phases 1 and 2 of this study showed that VR experience can improve learning in procedural tasks in lab-based STEM education, increases usability, and provide in-app data collection tools that can be used for evaluating performance. As the majority of IVR participants felt a strong sense of presence (as being real), they were more engaged and completed the tasks at a higher rate showing the positive impact of immersive VR.

Study 2a for comparing 2D, IVR, and DVR indicated that while 2D training in the form of text and video format is easier to use, the learning outcomes are significantly inferior (as seen in the post-knowledge test) compared to IVR and DVR, and other usability aspects show potential or clear value for VR applications.

4.6.1. Hypotheses

Hypothesis 1: Learning outcome will be positively impacted, confirmed.

The result shown in section 4.5.2 supported our first hypothesis that IVR and DVR training has a positive effect on learning outcomes as indicated by the higher score in the post-knowledge test. However, there was no significant difference between IVR and DVR in this area. It also did not support our expectation that the IVR group will score the highest as the DVR group achieved a better post-test score. Further analysis of the effect of gender (as covariance) on the difference between pre-post knowledge tests indicated no significant impact on the learning outcome.

While the result shows the DVR group is spending less time in the simulation, it did not necessarily mean it was more efficient or accurate in completing the tasks. On the contrary, the rate of incomplete tasks is higher in DVR which indicates that in an equal condition more IVR users are completing the experiments with success.

Hypothesis 2: 2D/ video will rank higher than IVR in usability, partially confirmed.

Our data analysis from the usability survey on seven areas of ease of user, memorability, learnability, pleasantness, clarity, visualization, and overall satisfaction, partially confirms our second hypostasis. The 2D condition ranked higher only in the first area of "ease of use" compared to IVR and DVR, and scored lower in every other area, although the difference was not always significant. This indicates that the 2D group was not as satisfied with the content presentation/delivery method as the VR group.

Through an open-ended question on how they felt about the experience, participants in IVR and DVR frequently commented on the "realism" of the environment and how "fun", "engaging" and "game-like" the experience was which indicates the high level of engagement and attraction of the experience. No participant in the 2D group added a similar comment or answer to the open-ended question. As such we conclude that participants had a more positive experience and feelings about learning in immersive VR than learning from a 2D method.

Hypothesis 3: DVR will rank higher than IVR in efficiency and accuracy, not confirmed.

Our data analysis (using ANOVA) did not show a significant difference between IVR and DVR environments on task efficiency (time to complete the tasks), which would contradict our hypothesis #3. This was particularly noticeable since several participants in the DVR group had commented about the difficulties to complete certain tasks such as picking up and rotating objects in the usability survey.

Further investigation within the in-app data log revealed that only one-third of the DVR participants have completed both training sections and one testing area of the experiment which justifies the lower number of generated data points within a similar period for IVR. A gender analysis on the task completion topic revealed that a higher number of female participants in IVR completed the task and only a quarter of the female participants in DVR completed similar tasks. We speculate that limited proficiency in using a mouse and keyboard combination can be contributed to a lack of experience in computer gaming since several computer games already apply such a combination.

A similar pattern was observed for the participants who came from the 2D group and were randomly assigned for testing in IVR and DVR. Those assigned to IVR performed better and completed more tasks compared to their DVR counterparts.

The accuracy of the task was mostly measured based on the collision detection records within the in-app data and a limited number of observations due to Covid-19 restrictions. From the observed and analyzed data we concluded that the IVR group has performed better and faster with more accuracy compared to the DVR group in completing the tasks supporting our 3^{rd} hypothesis.

For the difference in the results of pre-and post-knowledge test scores and in-app data, we can deduce that IVR is more impactful in engaging participants and producing a better learning outcome as shown in the data analysis. The result suggests that participants who viewed an immersive VR procedure reported significantly higher interest, and satisfaction and gained higher scores. The DVR group completion rate was low because interacting with virtual elements in DVR was not as smooth and easy with the combination of mouse and keyboard to complete the tasks efficiently. For 2D groups, it was easy to use, however, the post-test score was the worst among the three groups suggesting that lack of complete training (after they were randomly assigned to DVR or IVR) can be a factor in their lower post-test score or the current method using 2D content is not as effective as immersive VR to improve learning.

The in-app data collector proved to be a valuable source of information that gave us a fuller picture of the user's journey within the VR environment through each interaction (whether by the handheld controller in IVR or the keyboard stroke in DVR); the navigation, the path they take to complete a task and by observing how users explore the environment and solve the problems we gained further insights using this data.

The knowledge test does not provide such a comprehensive picture of the process and may not reflect the accurate learning outcome in comparison. While in this research we try to show the potential of in-app data collection in VR, the exact methods, and algorithms for collecting process metrics and performance evaluation based on them are beyond the scope of this research and are currently the subject of another Ph.D. thesis collaborating in the design of the data collection for Study 2a.

Hypothesis 4: IVR will create a strong sense of presence, confirmed.

As shown in section 4.5.5, the SUS survey revealed that the majority of IVR participants reported a strong sense of “being there” in the virtual space as we expected based on hypothesis #4. The overall results suggest a correlation between the feeling of “being” in the virtual environment and the similarity to the “real” world, as an attraction and encouraging factor to stay longer in the virtual environment and complete the tasks.

This feeling could have positively reflected on the learning outcome as the higher task completion rate and higher post-knowledge test scores.

SUS questionnaire would not be valid for DVR users since those participants were mostly using a different desktop or laptop screens and certain immersive features such as haptic feedback and surround sounds and 360 degrees rendered view was not available to them. Some of the participants in DVR agreed that the realistic lab environment game-line environment of the experience was noticeable and different from any of their prior learning experiences. The SUS survey did not apply to the 2D group.

Based on the above discussion we can argue that the answer to Q1 of this study is affirmative. The evidence suggests that teaching science lab procedures compared to the current 2D method is more effective in VR. To answer the second part of the Q2, we showed that while DVR and IVR share similar outcomes in certain areas, IVR provides a higher rate of engagement, performance, and efficiency and improved learning outcomes.

We also answered Q3 by designing and implementing an in-app data collection system for the VR lab that can be easily replicated and re-used in similar studies without the need for external APIs. The collected data provides a comprehensive view of the users' interaction with the environment and provides a rich source of data for researchers to investigate. Data collected from a knowledge test cannot compete with this value.

4.6.2. Limitations

The main limitations of the current study were related to hardware, software, and logistics issues. The installation process of the VR app on a new headset was challenging particularly for users with no experience with sideloading (i.e. installing an application from a source other than the official Meta /Oculus store). This step was needed for our users to download and install the VR app on their headsets to be able to run the experiment. We solved this issue by making sure that all the headsets were delivered to the participant with a pre-installed application. The issue of setting up the headset, guardian area, navigating, and locating the VR app was also solved by an extra information sheet (user manual) to enable participants to run the experiment.

As many aspects of this study including delivering headsets to IVR participants, data collection using surveys, pre-post knowledge tests, in-app data and even transferring the actual test app were completed remotely, the pandemic-related restrictions did impact our timeline.

The practical test in Study 2 was expected to be done in an actual lab after the ease of COVID restrictions. Replacing it with VR gave an advantage to IVR and DVR groups. During the COVID pandemic, recruiting interested participants was also a challenge.

While Study 2a provided some insights into the learning process in VR, the experience was mostly related to general interactions and not actual Chemistry content. To better evaluate the learning of STEM concepts, we designed Study 2b which included more clear content related to a Chemistry lab.

Chapter 5. Study 2b – Learning

In this chapter, we report on the process and findings of Study 2b which is focused on the effect of using VR learning outcomes and the interaction technique in IVR. Although the previous study provided some valuable insight into the effect of VR on learning outcomes, we identified the need to further focus on assessing the effect of Immersive VR and interaction techniques using curriculum-specific content as it was highlighted in the literature as a considerable gap.

The results of this study are being prepared for publication in the Journal of Education and Technology.

5.1. Overview

As mentioned in the previous chapters, despite advancements in Virtual Reality technology, the review of the literature shows several gaps[18] in research on how immersive virtual environments and interactions impact the learning process. The lack of curriculum-specific experiments in the current VR studies was identified as a significant gap [53]. Therefore, this study (Study 2b) was designed to address this gap by assessing the effect of authentic visualization and interaction types on learning science. A use case scenario of "Sp, Sp2, Sp3 orbital hybridization" in chemistry education is selected to create this experiment and to collect data for analysis.

While the study 2a was focused on the practicality, interactions, and usability of VR in education, study 2b is more focused on the interaction and learning effect of VR by using relevant and curriculum-based content and measuring user knowledge using standard methods of the knowledge test, survey, and in-app data.

The purpose of this study was to investigate and assess the effect of using immersive virtual reality (VR) and authentic three-dimensional (3D) visualization on learning compared to text-based/two-dimensional (2D) learning material. We also investigated the effect of interaction types (within the immersive VR environment), on user experience (UX) and learning outcomes.

18

-Learning theories
-Lack of fidelity/realism
-Lack of curriculum-specific content
-Assessment systems
-Accessibility/Inclusive design

While other forms of this study such as presenting the content on a tablet or using the actual ball-stick models could be alternative methods to use, they were beyond the scope of our study. The former would require producing another version of the experiment for the tablet platform (with touch interactions) and the latter would require access to several physical models to be distributed among participants during the pandemic.

In this study we aimed to answer the following research questions:

- *RQ3: Can using VR and authentic visualization (using ball-stick models) improve learning compared to two-dimensional (2D/Text-based) content?*
- *RQ4: Can flexible interaction type impact efficiency, accuracy, and learning outcome, if it enhances or hinders user experience (UX)?*

The result of this study shall complement the previous study conducted by the research team to develop a VR framework for STEM education. Our proposed experiment was designed to show a chemistry classroom with separate desks where each task will be conducted (Figure 29).

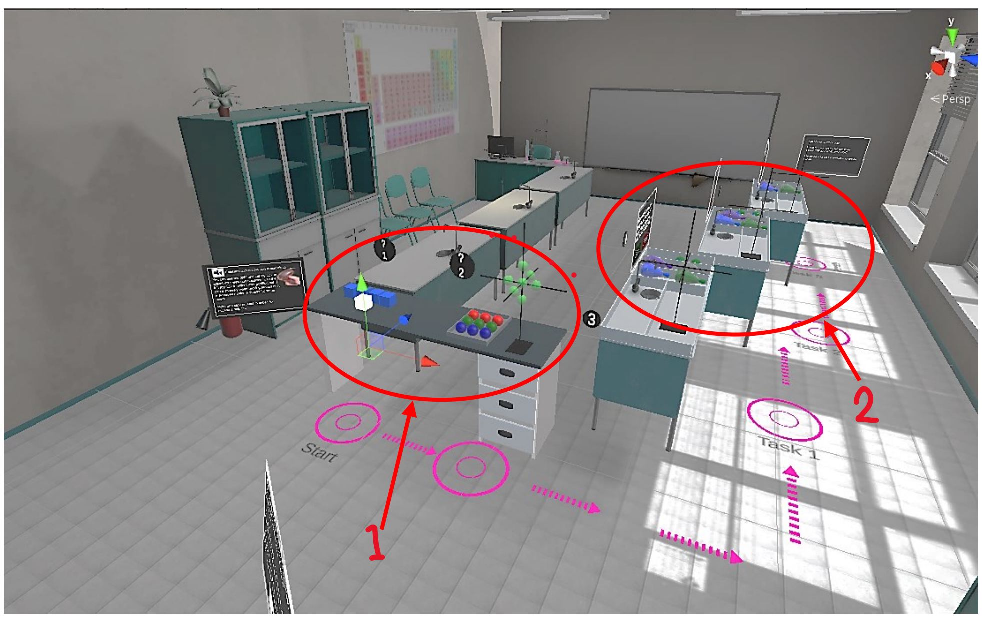


Figure 29.IVR experiment areas

The VR environment built for this study is a variation of the chemistry lab used in the previous study. The only difference was the VR room resembled a chemistry classroom rather than a real laboratory.

The environment change includes combining the tutorial area and main activity area to avoid the complexity of the scene transition. Since the selected curriculum-based activity was based on using ball-stick models (to build molecule structure), this modified environment was seen as more suitable for the designed tasks which were selected based on course instructors' decision. Also, as previously seen in Table 3, these tasks follow certain classifications related to VR affordances. These include objected selection (locating, verifying,), manipulation, (grasping and docking), and traveling (locomotion).

Also considering the fact the topic was being taught in the first year of chemistry education, we anticipated the possibility of using this environment for potential grade 12 high-school students who are interested to pursue education in a science-related field such as chemistry.

Objective data including a score from the knowledge test for all groups (Appendix II, B), and in-app data for both VR groups including task completion rate, efficiency, accuracy, and usability data (Appendix II, C) were collected. Subjective data included an open-ended question. We also had the opportunity for limited observation (during the ease of COVID restriction) to observe participants' activities and take notes (Table 8).

Table 8. Research Questions and Variables

Dependent Variable	RQ #	Type	Instrument
Learning outcome (SPS hybridization)	1	Objective	Knowledge test Observation
Usability	2	Subjective	Survey
Task completion (Grasp, Find, verify, travel, apply)	1	Objective	In-app data
Efficiency and accuracy	2	Objective	In-app data

5.2. Participants

Participants were undergraduate university/college students (preferably but not limited to first-year Chemistry or other science fields). The inclusion criteria were that participants should be able to use the Internet and be willing to use a VR headset (no previous experience was required). This study was conducted remotely (at participants' homes), with a total of thirty (30) participants divided into three groups:

- Group A, 10 participants (text-based group)
- Group B1,10 participants (IVR, interaction type1[19])

[19] Participants use the natural hand gestures and body movement to select, grasp, place/dock 3D objects.

- Group B2, 10 participants (IVR, interaction type2[20])

Ethics approval was received from the Carleton University Ethics committee. Participants received a $10 gift card as a token of appreciation for their participation in the study. In total, 30 users participated in this study, 14 female, 15 male, and 1 undisclosed gender.

5.3. Material

All hardware and software used in this study were the same as detailed in section 3.2.2. We also increased the number of headsets to 10 units (versions 1 and 2) to have a pre-installed application ready for users. Similarly, Survey monkey (surveymonkey.com) was used to create online surveys made available to participants to complete after the VR experience. The knowledge test and usability survey are similar to what was described in section 3.2 with needed changes to reflect the questions related to the current study (see Appendix II). These questions included general demographic information, handedness, and participants' experience with VR. The knowledge test comprised three questions with two components each. Participants were asked to draw orbital and Lewis structures (see next section for details) and identify the correct angles for three chemical compounds: Beryllium chloride (BeCl2), Boron trifluoride (BF3), and Ammonium (NH4). The score from the knowledge test would provide objective data to evaluate learning outcomes. To collect subjective data, our usability survey included questions on ease of use, memorability, learnability, pleasantness, clarity, visualization, and overall satisfaction to be rated on a scale of 1-5 (1 being very negative and 5 being very positive).

5.4. Procedure

A tutorial section was designed and developed to help VR users learn how to use handheld controllers that are used to interact with objects and the environment. Upon the start of the VR experience, a code (4 digits ID) would be automatically generated and assigned to each participant. They would use this code to complete the post-experiment knowledge test and survey.

For this study we defined three hypotheses to test:

- **Hypothesis 1:** Group A (2D), a knowledge test score will be lower compared to Group B1 and B2

[20] Participants use a laser beam (ray-cast), to interact with 2D buttons and to select and trigger the action to place/dock 3D objects.

- **Hypothesis 2:** Group B2 performs better in terms of efficiency and accuracy compared to the other 2 groups.
- **Hypothesis 3:** Usability test scores for group B2 will show higher rates for ease of use and overall satisfaction.

Each group was required to go through a specific procedure to complete the experiment described below:

5.4.1. Tasks for Group A (Text-based)

The participant was given a two-page of text/2D image content. They were expected to take 20-30 minutes to read the content. The goal was to see if participants can remember the proper angles for each type of hybridization. After the reading session, they completed a knowledge test for what they read. They were asked to complete the VR experience using a headset and then complete an online survey. This was only to allow this group to experience VR and discuss how they felt comparing text-based content with VR.

User journey for Group A:

1. Read the two-page lesson content prepared for them from the textbook related to the hybridization topic (appendix II, D)
2. Complete a knowledge test
3. Complete the VR experiment Interaction type 1 detailed in the next section.
4. Answer an open-ended question about how they felt in VR and if they thought it would enhance their learning if it had been made available to them before taking the knowledge test.
5. Complete a usability survey.

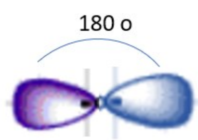
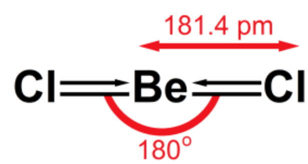
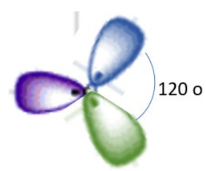
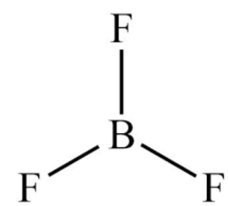
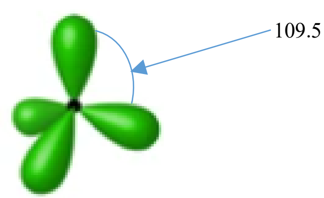
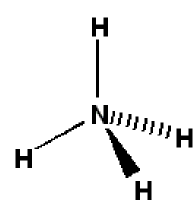
5.4.2. Tasks for VR Groups

VR groups were given 3D models (Figure 30) on a table in a virtual lab and were asked to arrange molecules with correct angles based on the information presented to them.

Figure 30.3D models representing the molecules available in the VR lab

Table 9 shows each task along with the expected correct answers for the post-experiment knowledge test. The goal of these tasks is to help participants learn the same course material and use proper angles for placing each 3D object on the stand. (task1 requires 180-degree placement, task2 requires 120-degree placement and task3 requires 109-degree placement). The snapping feature is enabled on each stand for multiple axes to challenge participants to reinforce the correct angle.

Table 9. VR Group Tasks and Expected Answers

Task	Correct Answer in 3D (Orbital Structure)[21]	Correct Answer in 3D (Lewis Structure)[22]
1-Build SP hybrid Orbitals and Lewis structures using BeCl2 as an example (linear).		
2: Build Sp2 Hybrid Orbitals and Lewis structures using BF3 (trigonal planar)		
3: Build Sp3 Hybrid Orbitals and Lewis structures using NH4 (tetrahedral)		

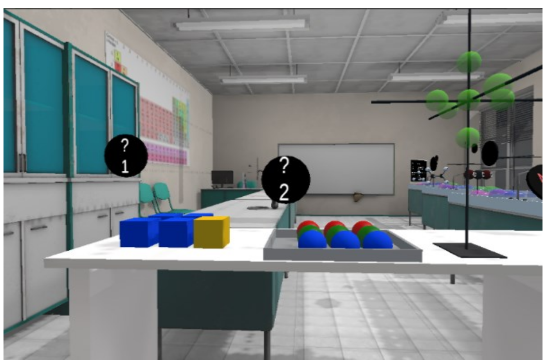
During our study 2a, we noticed several usability issues after converting immersive VR (IVR) interactions to Desktop VR(DVR) using One-to-one mapping. Participants in the DVR environment found the experience engaging and fun but had difficulty interacting with objects and environments using a keyboard and mouse. As a result, several participants did not complete the experiment.

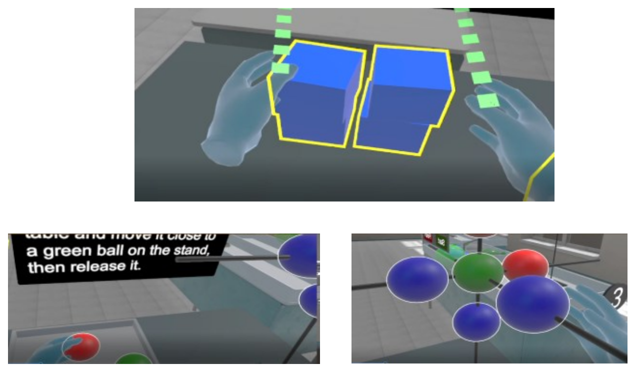
[21] An orbital is a three-dimensional description of the most likely location of an electron around an atom.
[22] A Lewis Structure is a very simplified representation of the valence shell electrons in a molecule.

For this study, we are proposing a combination of 2D interaction methods and 3D immersive visualization to test two different types of interactions in IVR and see if it would enhance user experience and usability.

5.4.3. Group B1-VR (Interaction Type 1)

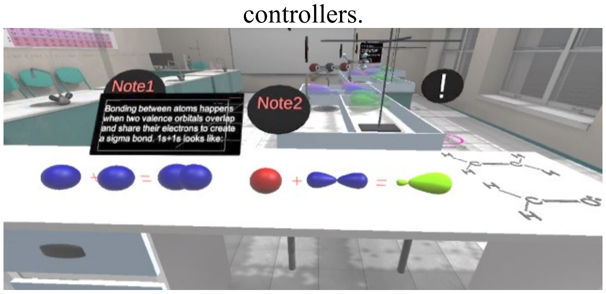
Participants in this group would start the VR experience first. They would complete a tutorial activity to learn how to use hand-held controllers and interact with objects. Participants were asked to follow sequential steps in VR to learn the same content as group A which was related to the topic of sp hybridization. This information was presented in VR through a series of interactive content, popup text, and voice. After completion of the VR experience, they were asked to complete the same knowledge test (as group A) and a usability survey. Figure 31(a-h) shows users' journey for Group B.



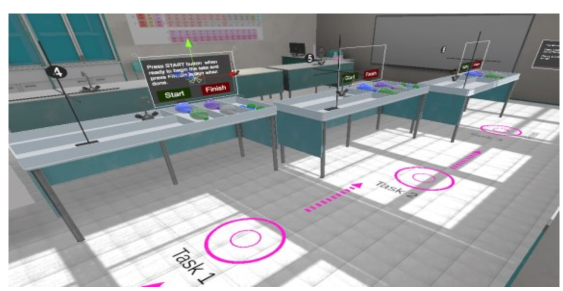


(a) User lands in the VR classroom and is asked to complete a tutorial as shown by Tasks 1 & 2. The goal here is to help participants learn how to use VR controllers.

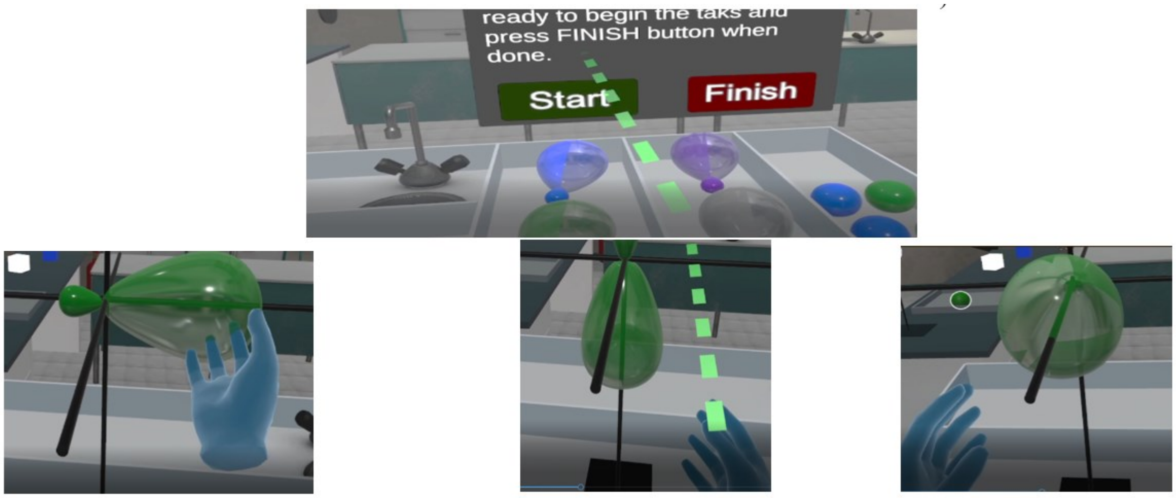
(b) Task 1 (top) is to stack cubes and task 2 (bottom) is to attach the balls to the stand



(c) User learns curriculum-based content in VR

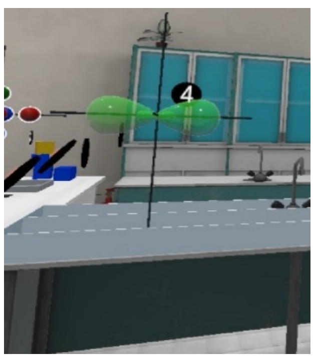


(d) User moves to the area for Tasks 1 to 3 (marked on the floor)

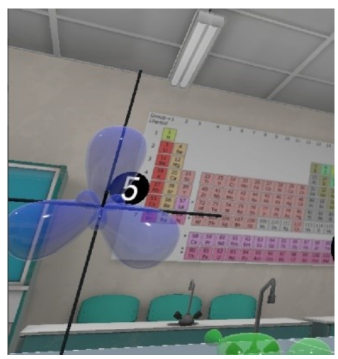


X Y Z

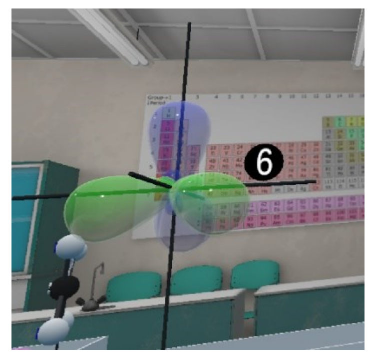
(e) User picks up 3D objects (top) and attaches the molecules to the correct axis (X, Y, or Z) based on the information provided to reinforce learning.



(f) User builds sp hybrid orbital (linear)



(g) User builds sp2 hybrid orbitals (trigonal)



(h) User builds sp3 hybrid orbitals (tetrahedral)

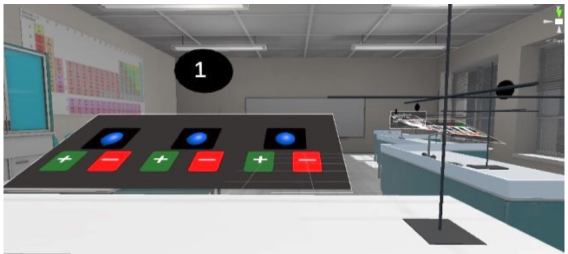
Figure 31.a-h: Users' journey for Group B

5.4.4. Group B2-VR (Interaction Type2)

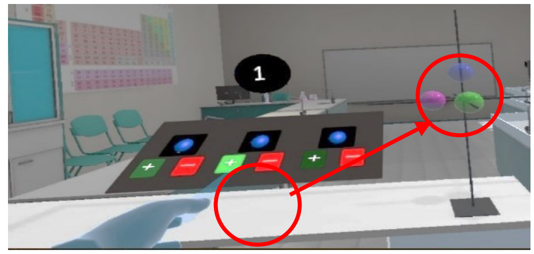
The procedures and tasks for this VR group were identical to VR group B except for the interaction technique. While the tasks are still objected selection, manipulation, and docking, the method of interaction is different from the previous VR group as participants use ray-cast to interact with 2D buttons in VR for automatic object placement.

This method of interaction was selected based on our observations in an earlier study comparing DVR and IVR and users' comments who felt this method would be a more efficient way to complete tasks and it is similar to using a mouse pointer and clicking. Therefore, it was added to test and measure the efficiency of ray-casting for interactions.

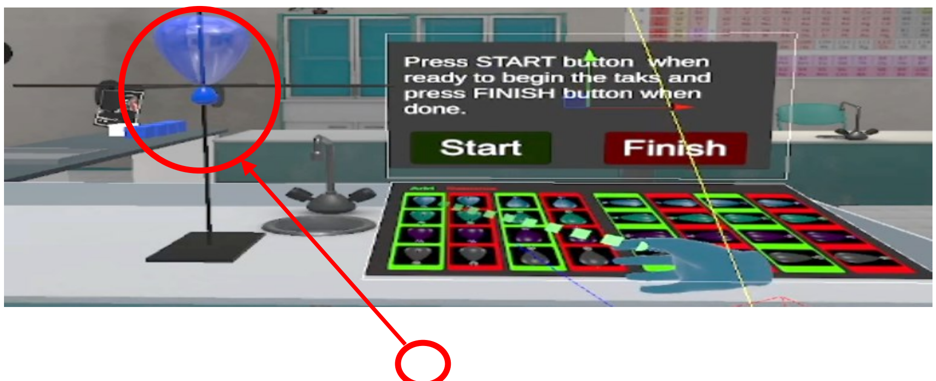
Participants in this group were expected to complete similar tutorial activities and main tasks as VR group B. However as explained earlier, the way they interact with objects is different. They use a raycaster to point to a 2D representation of a 3D object with the proper angle and use the trigger button on the controller to place the object on a pre-defined axis. Figure 32 (a-f) shows the user journey for this group (B2).



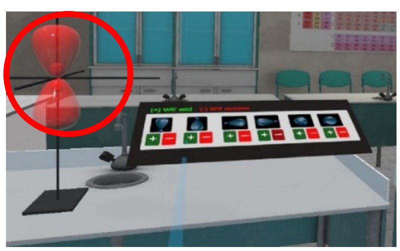
(a) User lands in the VR classroom and is asked to complete a tutorial activity.



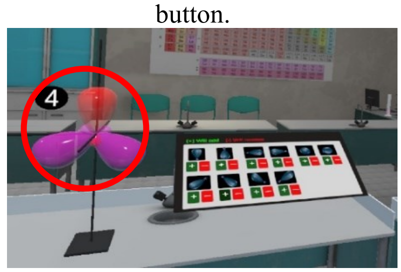
(b) User points to a 2D representation of a 3D object and uses the trigger button on the controller to place the object on a pre-defined axis.



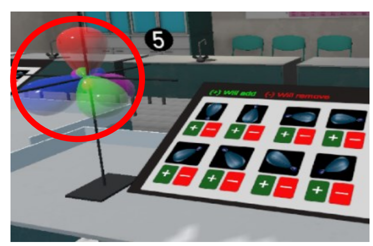
(c) A molecule is placed on the proper axis by pointing to a 2D representation of it using a ray-cast and trigger button.



(d) User builds sp hybrid orbital (linear)



(e) User builds sp2 hybrid orbitals (trigonal)



(f) User builds sp3 hybrid orbitals (tetrahedral)

Figure 32.a-f: Users' journey for group B2

5.5. Results

5.5.1. Demographic Information

The survey was completed by thirty (N=30) participants divided into three (3) groups. Group A consisted of individuals that completed a reading assignment (2 pages) and a knowledge evaluation before entering the VR environment. Group B1 was the first VR group to use the interaction type

1 technique (i.e. natural body and hand movement). Group B2 was the second VR group using interaction technique type 2 (i.e. ray-casting). As seen in Table 10 below 46.67 % of the participants were females and 50 % are males, with 6.67 % being left-handed vs 93.33 % being right-handed.

Table 10. Frequency of Demographic Factor Variables

Factor	Frequency
Handedness	
Left	2
Right	28
Gender	
Female	14
Male	15
Prefer not to say	1

H1: In Group A (text-based), a knowledge test score will be lower compared to Group B1 and B2.

Our hypothesis 1, was that group A is expected to achieve the lowest score on the knowledge test. This was due to the perceived complexity of the concepts and the lack of visualization features in text-based content vs what VR would be providing for the other two groups.

5.5.2. Orbital and Lewis Structure Difference for BeCl2

The knowledge test score differences for the compound BeCl2 were investigated between the three different groups in this section (Table 11). Given the non-normality of the data, a non-parametric approach was adopted to investigate the differences between groups. Non- parametric methods are used when data doesn't have a normal distribution or when we are using ordinal data such as the Likert scale and non-numeric labels.

Table 11. Knowledge Test Score Differences for Compound BeCl2

Group A		Group B1		Group B2	
Orbital	**Lewis**	**Orbital**	**Lewis**	**Orbital**	**Lewis**
1	0	0	1	1	1
0	1	1	1	1	1
0.5	0	1	0	1	1
0.5	0.5	1	1	1	0
0	0	1	1	1	0
0	0.5	1	0	1	0.5
0.5	0	1	1	1	0
0.5	0	1	0	1	0.5
0	1	1	0	1	0
1	0	1	1	1	1

A Kruskal-Wallis test (non-parametric alternative to one-way ANOVA) was carried out and based on the results, it was noted that there was a significant difference (**$P < 0.05$**) between the different groups for the orbital structure knowledge test. A post-hoc test was carried out to ascertain the differences that exist within the three different groups. Based on the result of the test, it was noted that there is a significant difference (**$P < 0.05$**) between "**Group A and Group B1**" and "**Group A and Group B2**"; whilst for "**Group B1 and Group B2**" there was no significant difference noted. As seen in Figure 33 below the distribution in Group B1 and Group B2 is higher than in Group A. This shows that in the knowledge test for the orbital structure of BeCl2, the VR groups performed better concerning the completion rate than the text-based group.

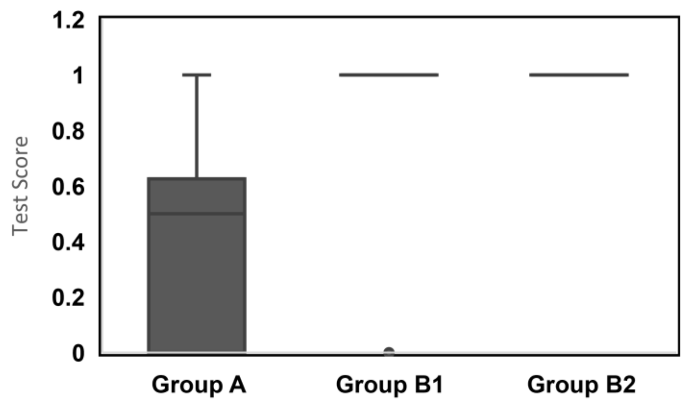


Figure 33.Box plot showing the distribution differences between groups for the Orbital structure knowledge test of the BeCl2

A Kruskal-Wallis test was also carried out to assess the differences in completion levels for the three different groups in the Lewis structure knowledge test. Based on the result it was noted that there is no significant difference (**$P > 0.05$**) in the distribution between the different groups for the Lewis structure knowledge test. This shows that all of the groups performed similarly concerning the completion level and accuracy on the Lewis structure knowledge test. Table 12 shows the summary of the completion rate in three groups for BeCl2.

5.5.3. Orbital and Lewis Structure Difference for BF3

In this section, the differences in knowledge test scores for BF3 between the three different groups were investigated. Given the non-normality of the data, a non-parametric approach was used to investigate group differences (Table 13). A Kruskal-Wallis one-way ANOVA was performed, and the results revealed a significant difference ($P < 0.05$) between the different groups for the orbital structure knowledge test. A post-hoc test was performed to determine the differences between the three different groups. Based on the result of the test, it was noted that there is a significant difference ($P < 0.05$) between "**Group A and Group B1**" and "**Group A and Group B2**"; while there was no significant difference between "**Group B1 and Group B2**".

Table 12. Descriptive Summary Completion Rate for BeCl2 Orbital and Lewis Structures

BeCl2	Completion rate	Group A	Group B1	Group B2
Orbital Structure	Not completed	4 (40%)	1 (10%)	0%
	Partly completed	4 (40%)	0.00%	0%
	Fully completed	2 (20%)	9 (90%)	10 (100%)
Lewis Structure	Not completed	6 (60.00%)	4 (40.00%)	4 (40.00%)
	Partly completed	2 (20.00%)	0.00%	2 (20.00%)
	Fully completed	2 (20.00%)	6 (60.00%)	4 (40.00%)

Table 13. Knowledge Test Score Differences for BF3 Compound

Group A		Group B1		Group B2	
Orbital	Lewis	Orbital	Lewis	Orbital	Lewis
0	1	1	0	1	1
0	1	1	1	1	1
0	0	1	0	1	1
0	0.5	0.5	0.5	0.5	1
0.5	0	1	1	1	0
0.5	1	0.5	1	1	0.5
0	1	1	1	1	0
0	0	1	1	1	1
0	0.5	1	0	1	1
0	0	1	0	0.5	1

As seen in Figure 34 the distribution in Group B1 and Group B2 is higher than in Group A. This shows that in the knowledge test for the orbital structure of BF3, the VR groups performed better concerning the completion rate than the text group.

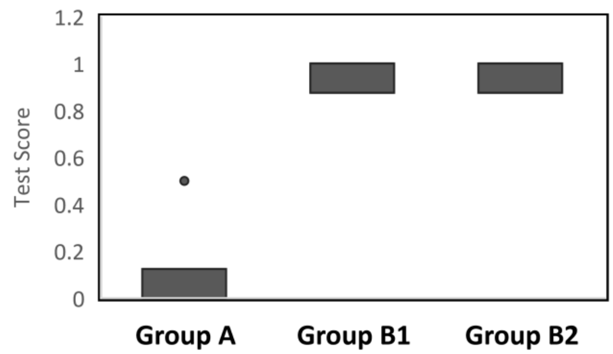


Figure 34.Box plot showing the distribution differences between groups for the Orbital structure knowledge test of the BF3)

For the Lewis structure knowledge test for the compound BF3, a Kruskal-Wallis test was also performed to determine the differences in the completion rate for the three different groups. According to the findings, there is no statistically significant difference ($P > 0.05$) in the distribution between the different groups for the Lewis structure knowledge test. This demonstrates that all of the groups performed similarly on the Lewis knowledge test for the compound BF3 in terms of completion rate as seen in Table 14.

Table 14. Descriptive Summary Completion Rate for for BF3 Compound

BF3	Completion Rate	Group A	Group B1	Group B2
Orbital Structure	Not completed	8 (80.00%)	0.00%	0.00%
	Partly completed	2 (20.00%)	2 (20.00%)	2 (20.00%)
	Fully completed	0.00%	8 (80.00%)	8 (80.00%)
Lewis Structure	Not completed	4 (40.00%)	4 (40.00%)	2 (20.00%)
	Partly completed	2 (20.00%)	1 (10.00%)	1 (10.00%)
	Fully completed	4 (40.00%)	5 (50.00%)	7 (70.00%)

5.5.4. Orbital and Lewis Structure Difference for NH4

In this section, the differences in knowledge test scores for NH4 between the three different groups were investigated (Table 15). Given the non-normality of the data, a non-parametric approach was used to investigate the differences between groups. A Kruskal-Wallis one-way ANOVA was performed, and the results revealed that there was no significant difference ($P > 0.05$) between the different groups for the orbital structure knowledge test for the compound NH4. Indicating that all of the groups performed similarly in terms of completion type on the Orbital structure knowledge test for the compound NH4.

Table 15. Knowledge Test Score Differences for NH4 Compound

Group A		Group B1		Group B2	
Orbital	Lewis	Orbital	Lewis	Orbital	Lewis
0	1	1	0	0.5	0.5
0	1	0.5	0	0.5	0
0	0.5	1	0	0.5	0
0.5	0	0.5	0.5	0	1
0.5	0	1	1	0.5	0
1	0	0	1	0.5	0
0	0	0.5	0	1	0.5
0.5	1	1	0	0	0
0	0.5	1	0.5	1	0
1	1	0	0	0	0.5

A Kruskal-Wallis test was also performed to determine the differences in completion rate between the three groups in the Lewis structure knowledge test for the compound NH4. According to the results, there is no significant difference ($P > 0.05$) in the distribution of the Lewis structure knowledge test between the different groups. This demonstrates that all of the groups performed similarly on the Lewis knowledge test for NH4 in terms of completion rate as seen in Table 16.

Table 16. Descriptive Summary Completion Rate for NH4 Compound

NH4	Completion Rate	Group A	Group B1	Group B2
Orbital Structure	Not completed	5 (50.00%)	2 (20.00%)	3 (30.00%)
	Partly completed	3 (30.00%)	3 (30.00%)	5 (50.00%)
	Fully completed	2 (20.00%)	5 (50.00%)	2 (20.00%)
Lewis Structure	Not completed	4 (40.00%)	6 (60.00%)	6 (60.00%)
	Partly completed	2 (20.00%)	2 (20.00%)	3 (30.00%)
	Fully completed	4 (40.00%)	2 (20.00%)	1 (10.00%)

H2: Group B2 performs better in terms of efficiency and accuracy.

Based on hypothesis 2, the second VR group (B2) using the ray-casting interaction technique would perform better compared to group B1. This was based on the observations made during a previous study on the use of ray-casts for selection tasks compared to natural body and hand movement using hand-held controllers to select and grasp objects in an immersive VR environment.

5.5.5. Efficiency Differences between Groups for the Tutorial Task

To determine the best statistical approach, the normality of the tutorial efficiency data was first investigated. Based on the findings, it was determined that the data is normally distributed, and thus a parametric statistical approach was preferred. Given that the goal of this section is to compare the differences in efficiency/duration between the three groups, a one-way ANOVA test was performed. According to the findings, there is a significant difference ($P < 0.05$) between the groups.

A Bonferroni correction post-hoc test was performed to determine the pairwise comparison between groups, and the results showed that there is no significant difference ($P > 0.05$) in the completion efficiency/duration between "Group A and Group B1." While there was a significant difference ($P < 0.05$) between "Group A and Group B2" and "Group B1 and Group B2".

As shown in Figure 35, Group B2 completed the tutorial task in less time than the other groups. As a result, Group B2 completed the tutorial task more quickly.

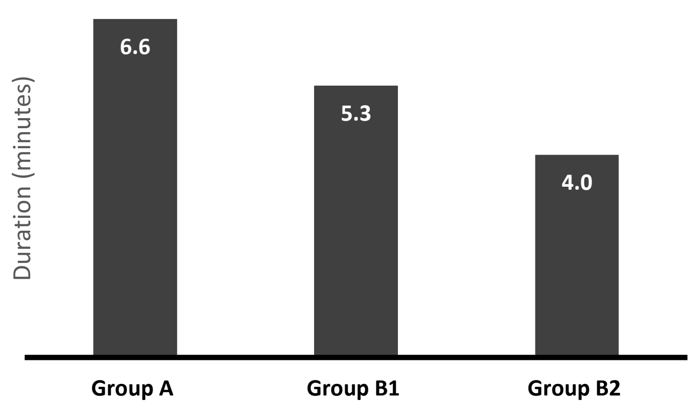
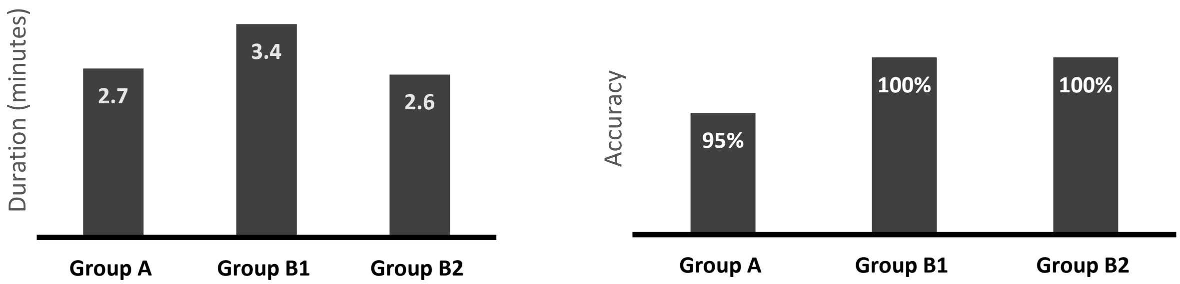


Figure 35.Bar chart showing the average duration differences for the three different groups

5.5.6. Efficiency and Accuracy Differences between Groups for Task 1 (BeCl2)

The normality of the efficiency and accuracy data for task 1 was first investigated to determine the most appropriate statistical approach. Based on the results of the result, it was noted that the data for both variables are normally distributed, thus a parametric statistical approach was favored. Given that the objective of this section is to evaluate the differences in duration and accuracy between the three different groups for task 1, a one-way ANOVA test was carried out. From the results, it was noted that there is no significant difference (**P > 0.05**) in duration between the groups for task 1 and also no significant difference in the (**P > 0.05**) accuracy percentage between the groups for task 1 (Figure 36).



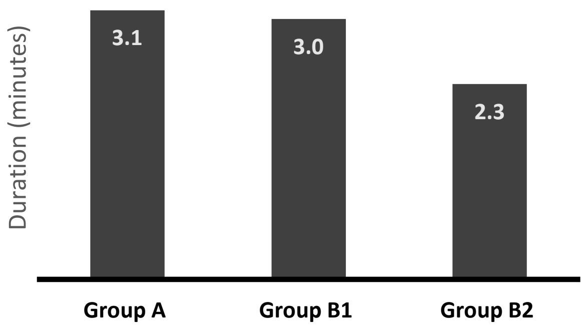
(a) (b)

Figure 36.Bar chart showing the duration differences

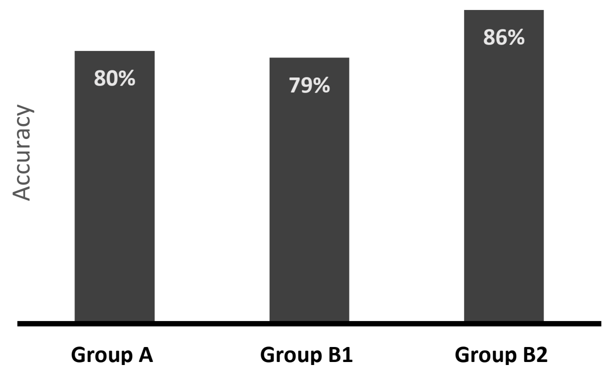
(a) average duration differences for the three different groups for task 1 and (b) average accuracy differences for the three different groups for task 1

5.5.7. Efficiency and Accuracy Differences between Groups for Task 2 (BF3)

Similar to task 1 the normality of the efficiency and accuracy data for this task was first investigated to determine the appropriate statistical approach. Based on the findings, it was determined that the data for both variables are normally distributed, thus a parametric statistical approach was preferred. Given that the goal of this section is to compare the differences in duration and accuracy between the three groups for task 2, a one-way ANOVA test was performed. According to the findings, although group B2 performs faster with more accuracy (Figure 37) there is no significant difference (P > 0.05) in task 2 duration between groups, as well as no significant difference (P > 0.05) in task 2 accuracy percentage between groups.



(a)



(b)

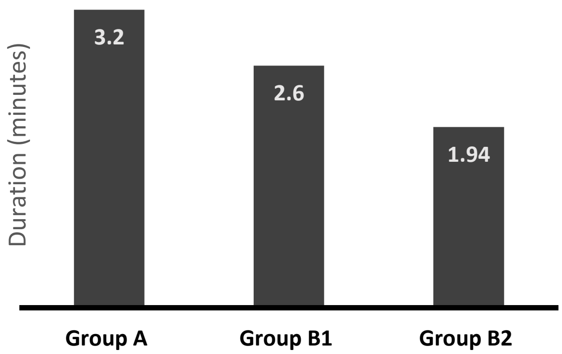
Figure 37.Bar chart showing the average duration differences

(a) average duration differences for the three different groups for task 2 and (b) average accuracy differences for the three different groups for task 2

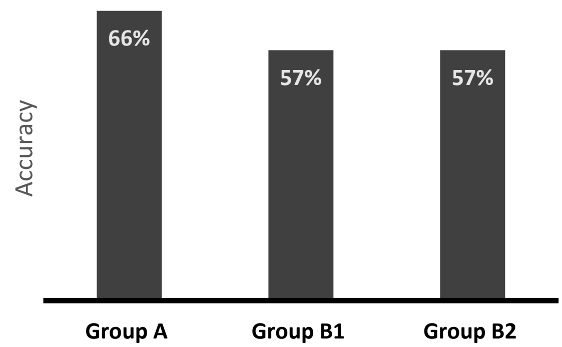
5.5.8. Efficiency and Accuracy Differences between Groups for Task 3 (NH4)

The normality of efficiency and accuracy data were first investigated for this task to determine a proper statistical approach. Based on the results of the result, it was noted that the data for both variables are normally distributed and thus, a parametric statistical approach was favored. Given that objective of this section is to evaluate the differences in duration and accuracy between the three different groups for task 3, a one-way ANOVA test was carried out. From the results, it was noted that there is a significant difference (**$P < 0.05$**) in duration between the groups for task 3. To ascertain where the differences occurred, a Bonferroni correction post-hoc test was carried and based on the result it was noted that there is no significant difference (**$P > 0.05$**) in the completion duration between "**Group A and Group B1**" and "**Group B1 and Group B2**" while a significant difference (**$P < 0.05$**) was noted between the "**Group A and Group B2**".

Since Group A (like Group B1), was using the interaction type 1 environment, this difference makes sense considering our hypothesis 2. As seen in Figure 38 below, Group B2 has a significantly shorter duration than group A on task 3 completion. Regarding the accuracy differences, a one-way ANOVA was carried out and it was noted that there is no significant difference in the (**$P > 0.05$**) accuracy percentage between the groups for task 3.



(a)



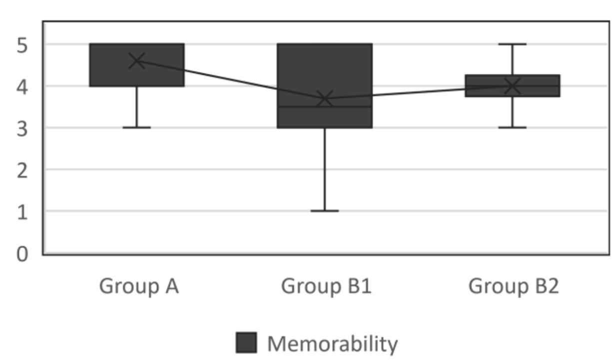
(b)

Figure 38.Bar chart showing the average duration differences

(a) average duration differences for the three different groups for task 3 and (b) average accuracy differences for the three different groups for task 3

5.5.9. Usability Results (Differences between Groups after VR Experience)

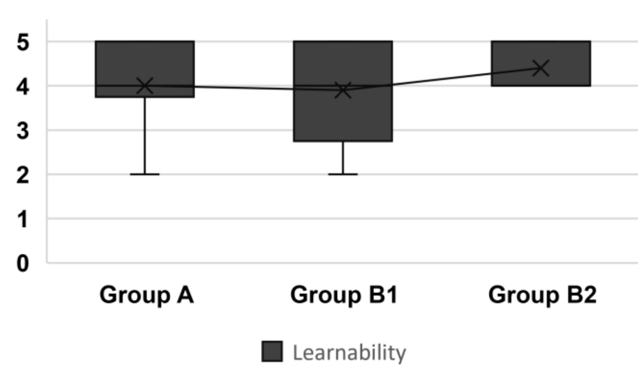
- **Memorability:** The memorability score differences between groups were investigated in the three. Given the non-normality of the data, a non-parametric approach was adopted to investigate the differences between groups. A Kruskal-Wallis one-way ANOVA was carried out and based on the results, it was noted that there was no significant difference in memorability (**P > 0.05**) between the different groups for the memorability score. Indicating that all of the groups had similar perceptions concerning the memorability of the VR experience (Figure 39).



(Likert scale: 1 very negative, 5 very positive)

Figure 39.Box plot showing the distribution of the memorability score

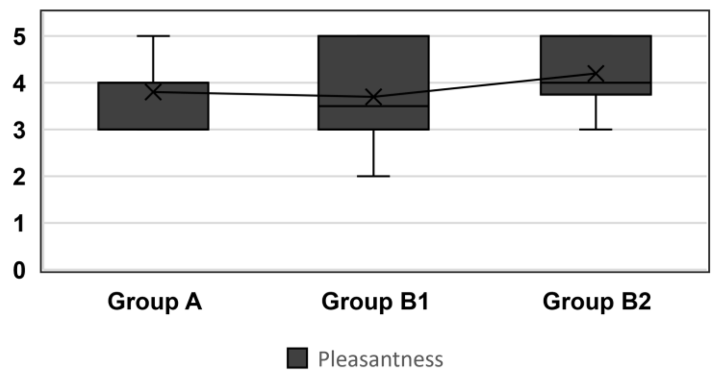
- **Learnability:** In this section, the differences in learnability scores between groups were investigated. Given the non-normality of the data, a non-parametric approach was used to investigate group differences. A Kruskal-Wallis one-way ANOVA was performed, and the results revealed that there was no significant difference in learnability score ($P > 0.05$) between the three groups. This indicates that all of the groups had a similar perception of the learnability of the VR experience (Figure 40).



(Likert scale: 1 very negative, 5 very positive)

Figure 40.Box plot showing the distribution of the Learnability score by groups.

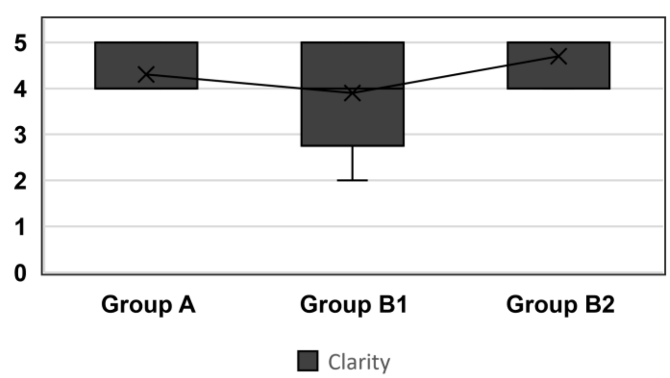
- **Pleasantness:** The pleasantness score differences between groups were investigated in this section in the three groups. Given the non-normality of the data, a non-parametric approach was used to investigate group differences. A Kruskal-Wallis one-way ANOVA was performed, and the results revealed that there was no significant difference in pleasantness ($P > 0.05$) between the groups. This indicates that all of the groups had a similar perception of the pleasantness of the VR experience (Figure 41).



(Likert scale of 1 very negative, 5 very positive)

Figure 41.Box plot showing the distribution of the Pleasantness score by groups.

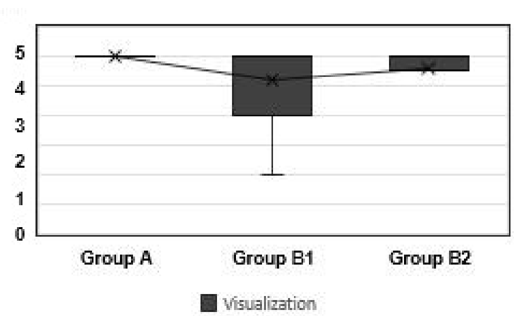
- **Clarity:** The differences in clarity scores between groups were investigated in this section. Given the non-normality of the data, a non-parametric approach was used to investigate group differences. A Kruskal-Wallis one-way ANOVA was performed, and the results revealed that there was no significant difference in clarity ($P > 0.05$) between the different groups. This indicates that all groups had a similar perception of the clarity of the VR experience (Figure 42).



(Likert scale of 1 very negative, 5 very positive)

Figure 42.Box plot showing the distribution of the clarity score by groups.

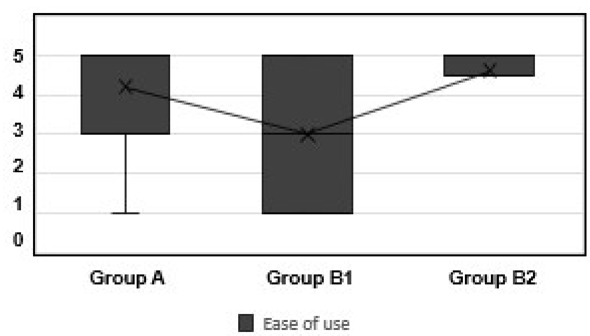
- **Visualization:** The visualization score differences between groups were investigated in the three groups. Given the non-normality of the data, a non-parametric approach was adopted to investigate the differences between groups. A Kruskal-Wallis one-way ANOVA was carried out and based on the results, it was noted that there was no significant difference in visualization (**$P > 0.05$**) between the different groups for the visualization score. It indicates that all of the groups had similar perceptions concerning the visualization of the VR experience (Figure 43).



(Likert scale of 1 very negative, 5 very positive)

Figure 43.Box plot showing the distribution of the Clarity score by groups.

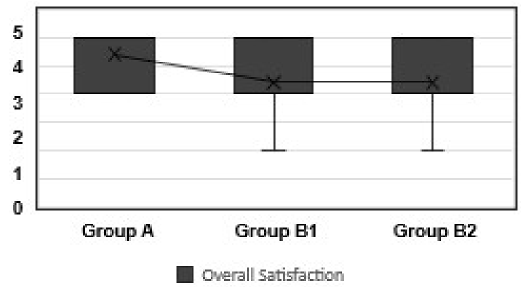
- **Ease of Use:** The Ease-of-Use score differences between groups were investigated in the three groups. Given the non-normality of the data, a non-parametric approach was adopted to investigate the differences between groups. A Kruskal-Wallis one-way ANOVA was carried out and based on the results, it was noted that there was no significant difference (**P > 0.05**) between the groups which indicates that all groups had similar perceptions of the Ease of Use of the VR experience (Figure 44).



(Likert scale of 1 very negative, 5 very positive)

Figure 44.Box plot showing the distribution of the Clarity score by groups.

- **Overall satisfaction:** The differences in overall satisfaction scores between groups were investigated in this section. Given the non-normality of the data, a non-parametric approach was used to investigate group differences. A Kruskal-Wallis one-way ANOVA was performed, and the results revealed that there was no significant difference in overall satisfaction ($P > 0.05$) between the various groups for the overall satisfaction score. This indicates that all the groups had a similar perception of overall satisfaction with the VR experience (Figure 45).



(Likert scale of 1 very negative, 5 very positive)

Figure 45.Box plot showing the distribution of the overall satisfaction score by groups

On the question of rating the statement "This VR visualization helps to understand the SPS concept better compared to what I read in the text" (Q9), the normality was assessed using the Shapiro-Wilk test. From the results of the test, it was noted that the data is normally distributed. (Table 17). Therefore, a parametric approach was adopted to assess the differences between the three groups for this question. The difference in the groups was analyzed using one-way ANOVA.

Table 17. Rating VR Experience on the Understanding of Concepts (Compared to Text)

Group A	Group B1	Group B2
5	3	4
3	4	4
5	5	5
5	4	3
4	5	3
5	5	4
5	5	4
4	4	5
3	4	4
4	4	5
Count: 10	Count: 10	Count: 10
Mean: 4.3	Mean: 4.3	Mean: 4.1
Median: 4.5	Median: 4	Median: 4
Standard Deviation: 0.823273	Standard Deviation: 0.674949	Standard Deviation: 0.737865

The f-ratio value is 0.23841, and the p-value is 0.78952. For this question, it was noted the result is not significant at $p < .05$, indicating that all groups had a similar perception of the effect of VR on their understanding of the concept compared to text.

Higher scores indicate that most of the participants agree with the statement that VR helps to understand the concept better than text. (Figure 46).

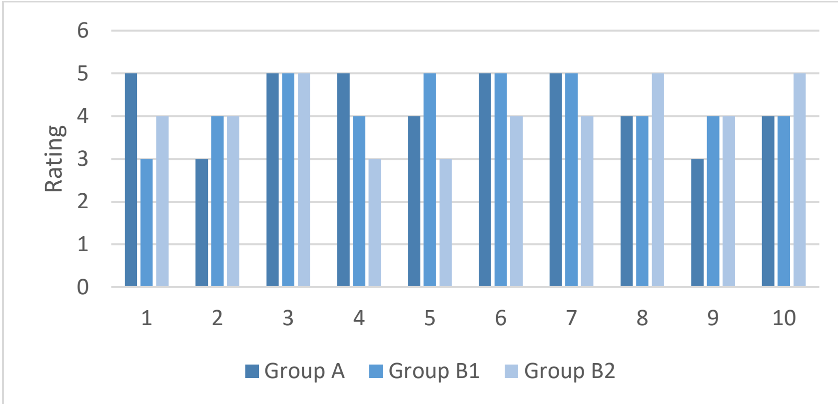


Figure 46.Ratings for the effect of VR on participants' understanding

And finally, to answer the question "Which of the following statements is agreeable to you?

a) My test score would improve if I had a chance to complete the VR experiment before taking the knowledge test"

b) Completing this VR experiment before the knowledge test would not help me to score better"

96.6% of participants (29 out of 30) believed that learning the course content through VR (before taking the knowledge test) would improve their test scores.

5.6. Discussion

5.6.1. Hypotheses

- **Hypothesis1:** The effect of VR on learning.

Considering test scores and analyzed data as detailed in the result section there were a few significant differences between Group A and B1 as well as group A and Group B2 concerning the Orbital structure and Lewis questions. These differences indicate that VR played a significant role in helping participants to answer orbital structure questions more accurately where 3D visualization was essential to understand and remember the concept due to the needed spatial skills.

However, there was no significant difference between groups B1 and B2 (VR groups) for most tasks regardless of the interaction type although the interaction type2 (ray-caster) performed slightly better in terms of efficiency and accuracy.

This confirms our H1 on Orbital structure questions, unlike in the case of Lewis structure questions between the three groups. This can be due to the insufficient interactive content developed within the VR for the Lewis structure topic (i.e. it was a more passive feature showing the Lewis structure users rather than making them interact or build the structure.) Figure 47 shows cases of presenting Lewis structure as passive content (non-interactable) in the virtual lab.

To answer the RQ3 we can deduce that using VR and authentic visualization has improved learning outcomes compared to 2D/text-based content.

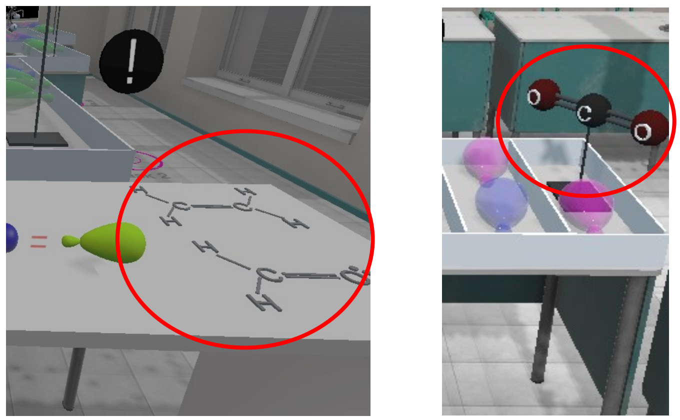


Figure 47.Lewis structures were presented as passive contents

It is worth mentioning that during our non-official user testing with other college students (in non-STEM fields), we noticed a similar effect. Meaning that participants could accurately remember and draw the orbital structure of each compound, confirming the impact of VR visualization in helping participants memorize and recall what they learned correctly.

- **Hypothesis 2:** For efficiency (in tutorial tasks):

There was a significant difference between the three groups to complete the tutorial task. Group B2 (interaction type 2, completed the task faster compared to the other groups). This confirms our H2 that Group B2 with the ray-casting interaction technique performs faster with higher accuracy.

  - **For efficiency and accuracy of Task 1 (Becl2):** Group B2 performs faster to complete that task (in 2.6 minutes vs Group B1 with 3.4 minutes and Group A with 2.7 Minutes). Group B1 and B2 also perform this task with 100% accuracy vs 95% for group A. However, it does not show a significant difference (statistically) which can be due to the small sample size. Given the opportunity to test more, we might see a larger difference.

- **For efficiency and accuracy of Task 2 (BF3):** Similar to takes 2, group B2 completed this task in a shorter time (2.3 minutes vs 3.0 minutes for Group B1 and 3.1 minutes for Group A), with more accuracy of 86% (vs 79% for group B2 and 80% for group A). However, it does not show a significant difference (statistically) which can be due to the small sample size.
- **For efficiency and accuracy of Task 3 (NH4):** There is a significant difference in the duration of task 3 between groups A and B2. It is worth mentioning that group A completed the VR experiment using an Interaction type1 similar to Group B1 (using body and hand movements). Overall results of comparing the three groups confirm our H2 that Group B2 performs faster with more accuracy showing that the ray-casting interaction technique provides better performance and accuracy in completing the tasks.

- **Hypothesis 3: Usability and overall satisfaction,** we expected that group B2 (using ray-casting) score a higher level however the result showed no significant difference between the groups using different interaction techniques.

From this result answering RQ4, we notice that interaction technique has a direct impact on the efficiency, accuracy, and learning outcome as seen in the results for group B2 meaning the remote interaction using ray-cast performed better compared to the body/hand movement.

An open-ended question about users' experience revealed that they felt the VR was “a great experience” that should be recommended and used in education. Some participants called it “amazing”, and a “very cool realistic lab”. One participant suggested including more practice questions in VR while another user commented on the level of detail and how “real” the lab looked. Some usability issues such as “teleportation” and occasional delay in teleporting were also mentioned.

5.6.2. Limitations

Updating software and using plugins became an issue that disabled parts of the project and had to be reconnected. Sharing the project source files with research collaborators was another issue due to the project size and complexity of Unity and Oculus Integration. It was solved by using the “Collaborate” feature of the Unity game engine.

Besides the usual restrictions due to COVID detailed in Chapter 4, Section 4.6.2, It was very challenging to recruit eligible users during this time. Since most of the users did not have a VR headset, researchers had to arrange to deliver a headset to each participant with all health-related precautions in mind.

Regarding the memorability score, the ideal condition would be to test users in a real lab scenario after two weeks however due to the pandemic restriction this part could not be achieved which was another limitation.

Studies 2a and 2b provided insights on general usability and learning in VR, but we were also interested in evaluating how VR can help with accessibility. As such, Study 3 was designed to start an initial investigation into the effect of VR in increasing inclusion and accessibility in STEM education.

Chapter 6. Study 3 – Accessibility

In this chapter, we report on the process and findings of Study 3 which was focused on the accessibility aspect of VR for wheelchair users and the effect of vertical movement as a feature to enhance accessibility. While the result of previous studies showed an improvement in the learning outcomes and usability of IVR, it lacked insight into the application of VR for users with accessibility needs. The main motivation for this study stems from the fact that addressing accessibility-related issues is not often a priority in the design phase and comes as an afterthought in the development phase.

Albeit a narrow focus, we intended to address this need and evaluate the effect of such a feature on learning outcomes for wheelchair users. The result of this study is approved to be published in IEEE AIVR 2022 with the title: "Improving Accessibility of Elevation Control in an Immersive Virtual Environment".

6.1. Overview

The progress made in both VR hardware and software during recent years has helped to make it more affordable and available. However, the VR experiences created during this period have not translated into it being more accessible for certain groups of users (i.e. wheelchair users)

Efforts for creating VR experiences that are accessible to a wider range of users (especially wheelchair users) seem to be too slow, as reviewed in Chapter 2, Section 2.5. For example, safe zones of movement for headset-based VR experiences can be small and uncomfortable for wheelchair users. Similarly, reaching higher or lower points in the VR environment to grab and manipulate objects can be difficult for these users. As a result, people with such disabilities are unable to use VR systems effectively, even though the programmable nature of VR and other digital environments promises more accessibility than the physical world. While our previous studies provided insights into the learning and interactions aspect of a VR space, they did not address the accessibility challenges of VR for wheelchair users.

In this study, we investigated how to facilitate vertical movement in a VR environment for wheelchair users. The specific questions for this study were:

- *RQ5: Can software controls for vertical movement provide improved accessibility for wheelchair users with task efficiency and completion rate similar to non-wheelchair users?*
- *RQ6: Can the improved accessibility (if any) improve learning?*

While various forms of customization of the environment could be considered, we focus on software controls to increase and decrease the user's virtual elevation. Vertical movement for wheelchair users was selected since it can be a major accessibility issue and has been a less considered topic in research studies. We hypothesize that this feature allows a simple yet effective way to use the same virtual environment as non-wheelchair users with minimum customization effort. The result of this study can contribute to creating more accessible VR experiences for wheelchair users. Increased accessibility can help this group of users to complete tasks more efficiently when compared to non-wheelchair users.

It is worth mentioning that without the accessibility feature allowing users to change their virtual height, wheelchair users in this study would not be able to complete certain tasks that would require raising or lowering the body to reach an object. The same access limitation exists in the real world for wheelchair users, where lab equipment can be out of reach, particularly in the labs that are not modified or enhanced for such users.

Objective data including a score from the usability survey (appendix III), and in-app data for both VR groups including task completion rate, efficiency, and accuracy were collected. Subjective data included an open-ended question. We also had the opportunity for limited observation (during the ease of COVID restriction) to observe participants' activities and take notes (Table 18).

Objective data:

- Score from survey
- In-app data collector for both VR groups: task completion, efficiency, accuracy

Subjective data:

- Open-ended question
- Limited observation data (if classes were held in person and we could observe participants).

Table 18. Research Questions and Variables

Dependent Variable	RQ #	Type	Instrument
Task completion (Grasp, Find, verify, travel, apply)	1	Objective	In-app data
Efficiency and accuracy	1	Objective	In-app data
Learning	2	Subjective	Post experiment survey Observation

6.2. Participants

Participants were undergraduate university or college students in Science or Engineering programs. The inclusion criteria were that participants should be able to use the Internet and be willing to use a VR headset (no previous experience was required).

Participants were assigned to three groups:

- Group A, 10 participants: non-wheelchair users with no accessibility feature available
- Group B, 10 participants (verifying group): non-wheelchair users with accessibility features available
- Group C, 10 participants: wheelchair users with accessibility features available

Participants were recruited based on a gender-balanced participants group. The verifying group was asked to use the accessibility feature for comparison with wheelchair users.

We had ten (10) participants in each group (Total N=30). Groups B and C could access to height changing feature by pressing the X button on the left-hand controller (Figure 48.) It featured two types of UI elements to help users adjust their height and increase or decrease access elevation. Using the Up/Down button would change the height in 20-centimeter increments, while the Slider feature would smoothly change the height with 2-5-centimeters increments that were controlled by a visible laser beam emitting from handheld controllers known as "ray-casting".

This study was approved by the institutional Ethics Board and was conducted online with downloaded programs for VR headsets. Participants received a $10 gift card as a token of appreciation for their participation in the study.

During the design phase, we consulted with actual wheelchair users and used their feedback to determine accessibility features and UI for this study.

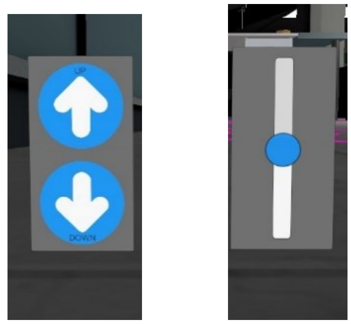


Figure 48.The "UP/Down" and "Slider" accessibility features

6.3. Material

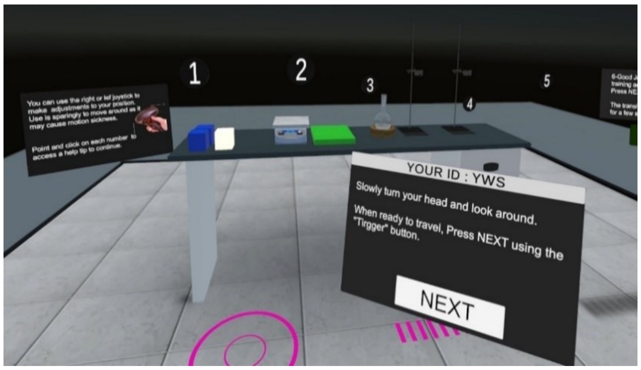
All hardware and software used in this study were the same as detailed in section 3.2.2. The VR experiment was developed using an updated version of Unity 2021, plus the latest oculus Integration Software Development Kit (SDK). All the objects and assets were either modeled using 3DS Max software (where each object was exported into FBX format and imported into Unity) or acquired through the Unity asset store. Online tools, surveys, and in-app data collection methods were identical to previous studies. Survey questions were modified based on the current study requirements (Appendix III). The post-experiment survey included general demographic information, handedness, and participants' experience with VR. It also featured three questions related to ease of use, pleasantness, and ease of learning for three types of tasks: 1) tasks that did not require raising or lowering the body, 2) tasks that required lowering the body 3) tasks that required raising the body rated by a Likert scale of 1-5 (detailed in the next section).

6.4. Procedure

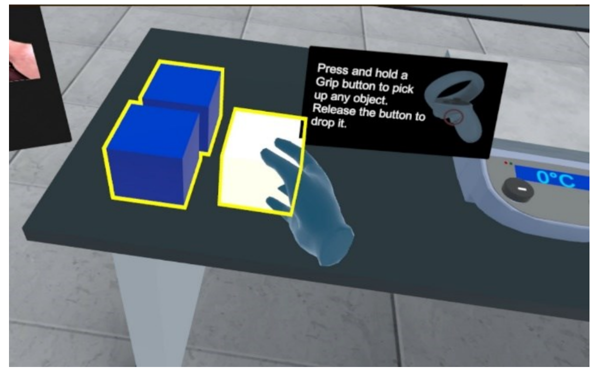
The experiment for each group has two parts: (1) completing a list of tasks (listed in the next section) and narrating to them in the VR environment, and (2) completing a post-experiment online survey. This VR environment has two separate areas. One for the basic tutorial and one for the main activity. The experiment was built based on real-world interactions for selecting and picking up objects in a virtual chemistry lab placed at different elevations (Figure 50). Participants were asked to complete the tasks by lowering and raising their hands/arms and body to reach objects.

Their interaction data was collected by the in-app data collector. After the completion of the VR experience, participants were asked to complete a usability survey based on 5 points Likert scale that included (i) how pleasant the activity was (ii) how easy the task was to perform (iii) how easy it was to learn the task.

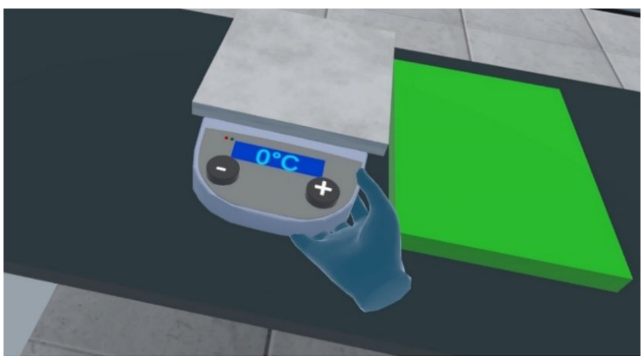
In this VR environment participants were expected to spend 20-30 minutes completing the required tasks detailed in the next section. The image sequence in Figure 49 (a-d) shows the users' journey within the tutorial area.



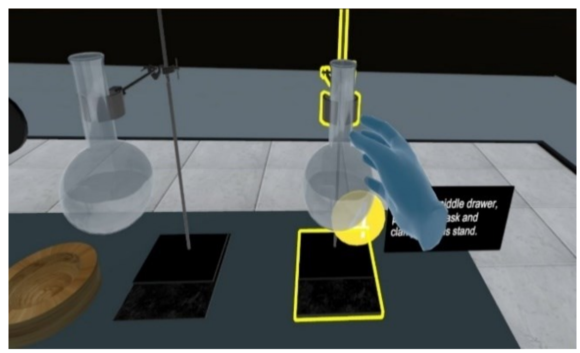
(a) User wears the VR headset and starts the VR experience, user is also assigned an automatic ID, and guided to the next steps by voice and text



(b) User learns simple interactions using controllers by stacking cubes



(c) User learns to interact with virtual lab equipment



(d) User learns to grasp, place and release objects

Figure 49.a-d: Users' journey within the tutorial area

For the highlighting and grasping tasks, some of the interactable objects were placed at different elevations (Figure 50). For example, a flask was placed on the top shelf requiring the user to reach up, or a piece of equipment was placed inside a drawer that required the user to lower the body to open the drawer and pick it up.

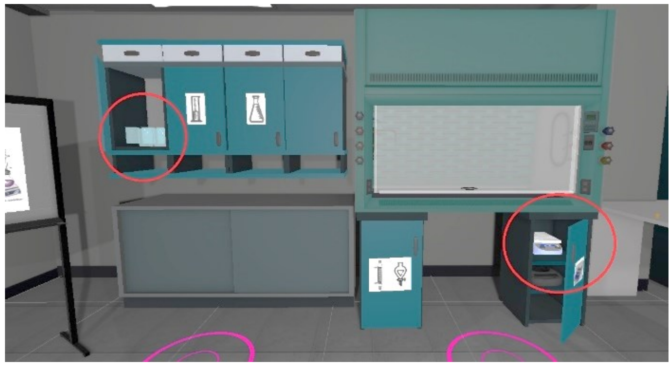
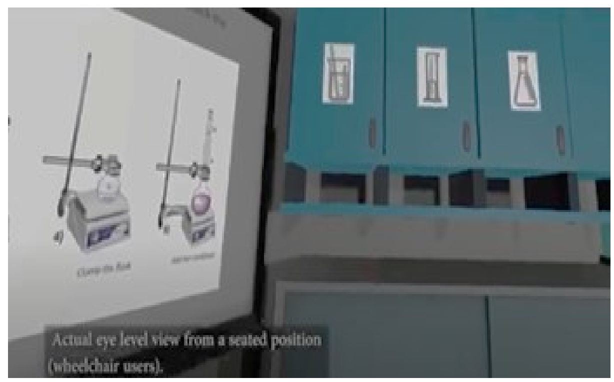


Figure 50.Lab equipment and props placed at different elevations

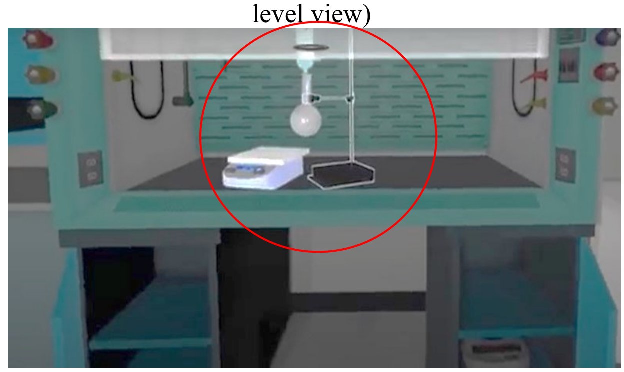
The main activity area required participants to follow a few instructions to set up a chemistry experiment using virtual equipment as shown in Figures 51 (a-d).

(a) User wears the VR headset and starts the VR experience



(b) User reads (and hears) the instructions to set up a chemistry experiment (This is a wheelchair user's eye-level view)

(c) Users will locate objects and set them up according to the instructions. Wheelchair users can use the accessibility feature to increase/decrease the height



(d) User completes the setup and places the equipment in a fume hood.

Figure 51.a-d: The main activity area

A video clip showing the application of accessibility features can be viewed here : https://drive.google.com/file/d/1TeCXAKqAKcOq7aaJK3xME3TpMwkk4DYK/view?usp=sharing

Participants in the VR lab would interact with objects and the environment. The tasks to complete included:

- Locating objects: By looking around the environment and locating objects in response to the task description provided to them in pop-up windows.
- Selecting: An interactable object would be highlighted using ray-cast (Figure 52).

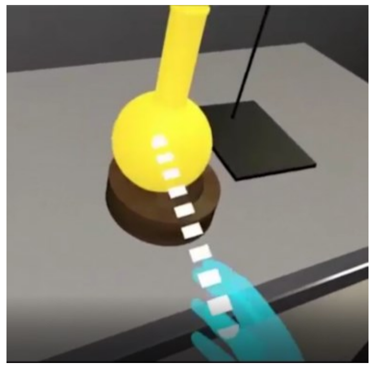


Figure 52.A highlighted round flask

The dotted line represents a ray-cast

- Grasping: virtually picking up an object using the handheld controller and its "Grip" button (Figure 53).

Figure 53.Grasping (picking up an object)

- Releasing the grip/mouse button could result in the virtual object falling, snapping to a predefined place, or being placed at any location chosen by the participant.
- Releasing/docking: After grasping an object, the user will move it to a target area and release it at that place. For example, attaching a flask to a tripod.
- Manipulating: This involves pressing a button to interact with menu items and dialog boxes.

- Traveling/teleporting: mainly for horizontal and/or precise movement by using the controller buttons for larger distances (Figure 54), or for physical movement such as walking inside the pre-define safe area (shorter distances) using the joystick movement to travel. Users were cautioned about using this feature as it could induce cybersickness in some users (3% of users in our case had a mild level of cybersickness).

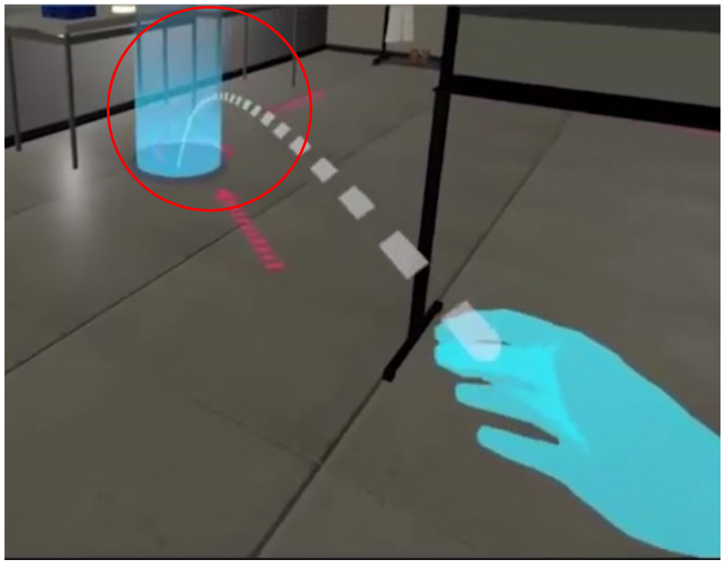


Figure 54.Teleporting to a target point using a VR controller

For this study we defined the following hypotheses to test:

Hypothesis 1. Accessibility features should improve learning, task completion rate, and accuracy for group C (wheelchair users).

Hypothesis 2. The overall results between group B (verifying group) and group C (wheelchair users) should be close in terms of task completion and accuracy.

Hypothesis 3. The slider accessibility feature will be preferred and performs better than the Up/Down accessibility feature.

6.5. Results

6.5.1. Demographic Information

In total, 30 responses were recorded in the survey, which was made up of three groups of equal participants. 53% were females and 47% were males. Most of the participants (83%) were right-handed. The average age of participants was 24.5, with the youngest participant being 16 and the oldest participant being 43. The participants in each group were asked if they had any prior experience with VR. The percentage and types of their VR experiences are as follows:

Group A:70%, Group B: 50%, Group C: 30%

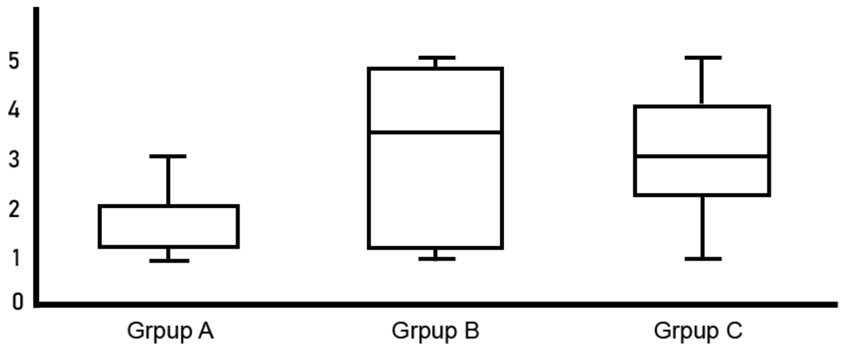
All three groups mentioned VR games, popular 360-degree VR videos, and similar forms of experiences using lower-end VR devices such as cardboard. A Chi-squares [118] test of independence was carried out to determine if there is a relationship between participant grouping and experience with VR. The results showed that both categories were independent of each other ($p > 0.05$), indicating that the fact that most experienced participants were in Group A did not mean it would stay the same if the population size increased.

6.5.2. Tasks with no Vertical Movement

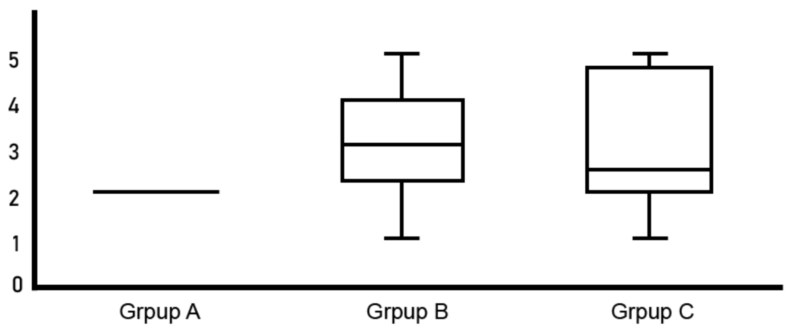
Tasks that didn't need vertical body movement were designed to establish a baseline which helped us understand if there are any differences in the three groups of participants. To ascertain the perception of participants on VR tasks that did not require raising or lowering of the body, we asked:

- (i) How pleasant was the activity?
- (ii) How easy was it to perform the task?
- (iii) How easy was it to learn the task?

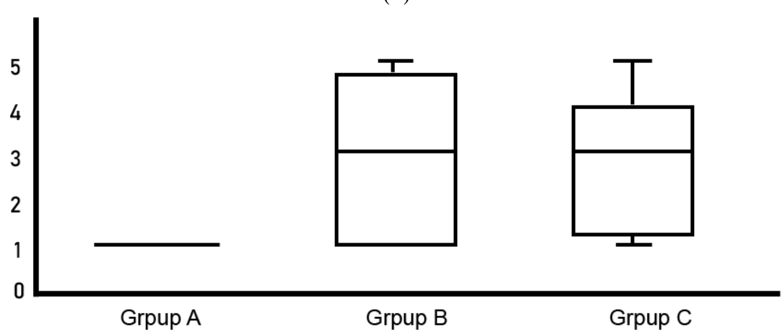
A Shapiro-Wilk normality test [119] was carried out to verify the parametric assumptions that the data are normally distributed ($p < 0.05$). Based on the results, it was noted that the data from the survey did not follow a normal distribution (Figure 55a-c). Thus, a non-parametric test was favored. Non- parametric methods are used when data doesn't have a normal distribution or when we are using ordinal data such as the Likert scale and non-numeric labels. Given the fact that we have three groups, and the aim is to assess the differences in each, the Kruskal-Wallis non-parametric test [120] was selected to help determine the distribution differences.



(a)



(b)



(c)

Figure 55.Distribution of responses

For tasks that did not require lowering or raising the body. (a) How pleasant, (b) How easy to perform, (c) How easy to learn

From the output of the analysis (Table 19), it was noted that for the question "How pleasant the activity was" and "How easy it was to perform the task", there is no significant difference in the distribution of responses ($p > 0.05$). However, for the question "How easy it was to learn the task", the distribution of the data was noted to be significantly different ($p < 0.05$) between the three different groups. Further inquiries were carried out to ascertain where these differences occur, using a post hoc test.

It was noted that Group A's responses were significantly different ($p < 0.05$) from Group B and C. While Group B and C were noted to have no significant difference in distribution. This indicates that Group A has significantly more individuals who agree with the statement that learning tasks that did not require body movement were easy.

Table 19. Summary of Tasks without Lowering/Raising Body

How pleasant the activity was (1 pleasant,5 unpleasant)			
Scale	Group A	Group B	Group C
1	16.67 %	9.68 %	10.13 %
2	66.67 %	6.45 %	20.25 %
3	16.67 %	9.68 %	18.99 %
4	-	25.81 %	25.32 %
5	-	48.39 %	25.32 %
How easy was the task to perform (1 easy, 5 difficult)			
1		3.13 %	6.67 %
2	85.71 %	12.50 %	20.00 %
3	14.29 %	28.13 %	10.00 %
4	-	25.00 %	13.33 %
5	-	31.25 %	50.00 %
How easy it was to learn the task (1 easy, 5 difficult)			
1	81.82 %	13.79 %	10.34 %
2	18.18 %	6.90 %	13.79 %
3	-	-	-
4	-	27.59 %	41.38 %
5	-	51.72 %	34.48

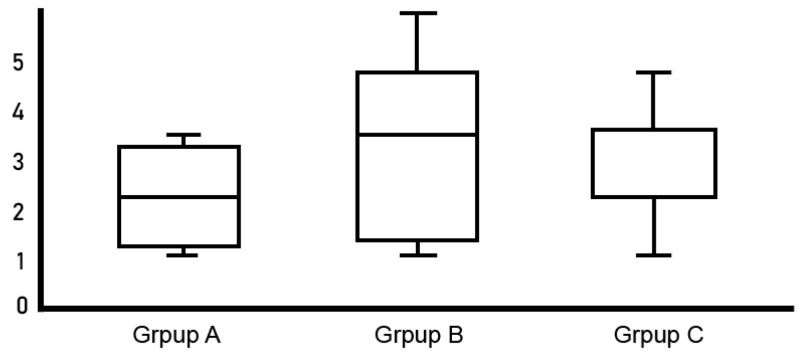
As the aim of this section was to assess the differences in the distribution of responses for the task that did not require lowering/raising of the body, a Mann- Whitney-U test was carried out. From the output, it was noted that there is no significant difference ($p > 0.05$) in the distribution of the responses.

6.5.3. Lowering the Body

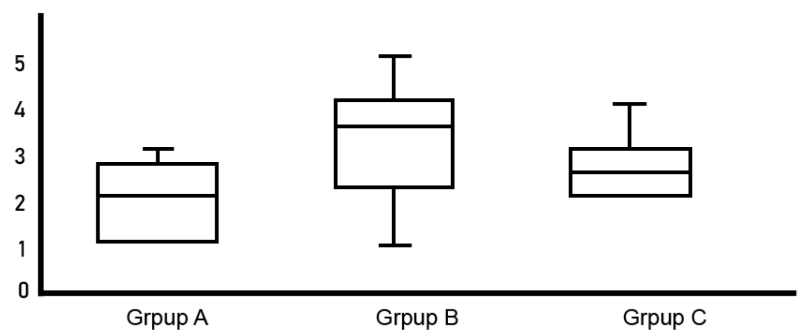
To ascertain the perception of participants on VR tasks that required lowering the body, we asked the same three questions as in section 6.5.2 (Figure 56a-c).

The normality of the data was assessed using the Shapiro-Wilk normality test, due to the small sample size. The results showed that the data is not normally distributed ($p < 0.05$), indicating that the responses have violated the assumptions for a parametric test, and therefore, Kruskal-Wallis non-parametric test was utilized to assess the group differences. This analysis helped to assess the mean rank differences.

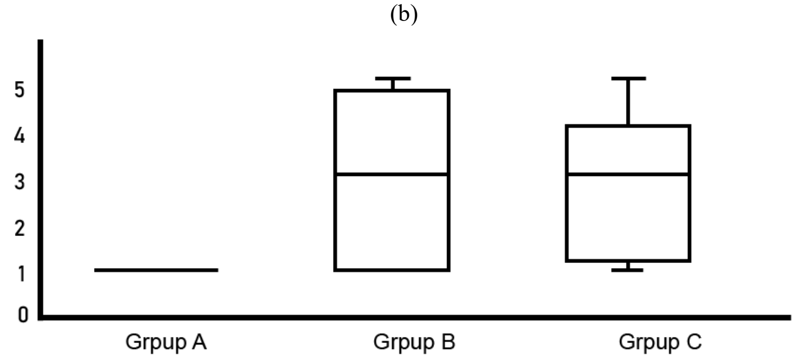
For the question "How pleasant the activity was" and "How easy it was to learn the task", there is no significant difference in the mean rank for the distribution of responses ($p > 0.05$). For the question of "How easy it was to perform the task", the distribution was noted to be significantly different between the groups. A post-hoc test was conducted, and the results indicated that Group C's responses were significantly different ($p < 0.05$) from A and B's. While groups A and B were noted to have no significant difference ($P > 0.05$) in distribution.



(a)



(b)



(c)

Figure 56.Distribution of responses

For tasks that required lowering the body. (a) How pleasant, (b) How easy to perform, (c) How easy to learn

As seen in Table 20 below, groups A and B have significantly more individuals who agree with the statement that learning tasks that require lowering body movement were easy.

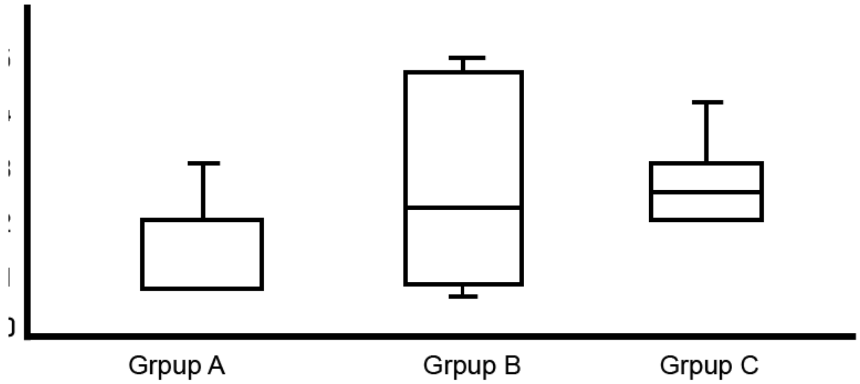
Table 20. Summary of Tasks with Lowering Body

How pleasant the activity was (1 pleasant,5 unpleasant)			
Scale	Group A	Group B	Group C
1	15.00 %	10.34 %	3.57 %
2	40.00 %	13.79 %	21.43 %
3	45.00 %	-	42.86 %
4	-	41.38 %	14.29 %
5	-	34.48 %	17.86 %
How easy was the task to perform (1 easy, 5 difficult)			
1	21.05 %	3.03 %	-
2	31.58 %	12.12 %	35.71 %
3	47.37 %	18.18 %	32.14 %
4	-	36.36 %	14.29 %
5	-	30.30 %	17.86 %
How easy it was to learn the task (1 easy, 5 difficult)			
1	66.67 %	14.81 %	25.00 %
2	33.33 %	7.41 %	20.00 %
3	-	11.11 %	30.00 %
4	-	29.63 %	-
5	-	37.04 %	25.00 %

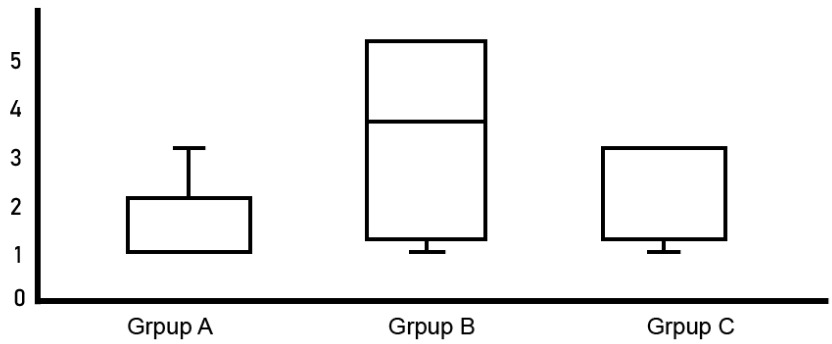
In this section, the differences in the mean rank of the responses for the task that required lowering of the body were assessed using a Mann-Whitney U test and provided the comparison of the two groups. From the result, it was noted that there is no critical distinction ($p > 0.05$) in the mean rank between experienced and non-experienced members.

6.5.4. Raising the Body

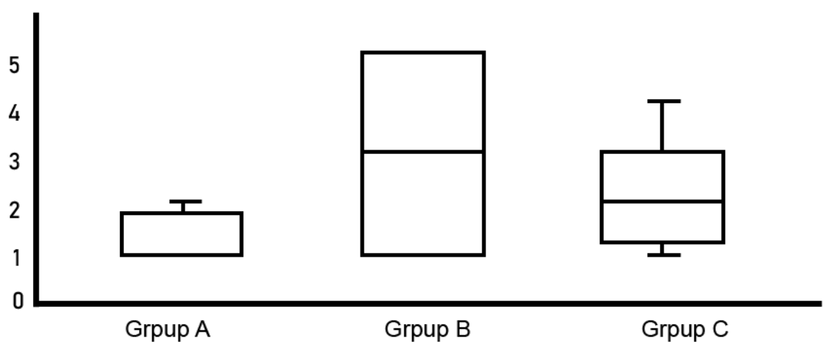
To ascertain the perception of participants on VR tasks that required raising the body, we asked the same three questions as in section 6.5.2 (Figure 57a-c).



(a)



(b)



(c)

Figure 57.Distribution of responses

For tasks requiring the raising of the body. (a) How pleasant, (b) How easy to perform, (c) How easy to learn

The results from the normality tests showed that the data is not normally distributed ($p < 0.05$), indicating that the responses have violated the assumptions for a parametric test. The Kruskal-Wallis non-parametric test was selected to analyze the data. This analysis helped to assess the mean rank differences.

From the output of the analysis (Table 21), it was recorded that for the question "How pleasant the activity was", "How easy it was to perform the task" and "How easy it was to learn the task", there is no significant difference in the mean rank for the distribution of responses ($p > 0.05$).

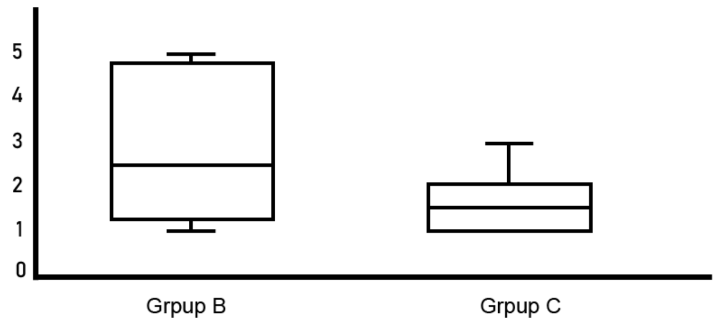
The differences in the mean rank of the responses for the task that required raising of the body were assessed using a Mann- Whitney U test to provide comparisons between the two groups. From the result, it was noted that there is no critical distinction ($p > 0.05$) in the mean rank between experienced and non-experienced members.

Table 21. Summary of Tasks with Raising Body

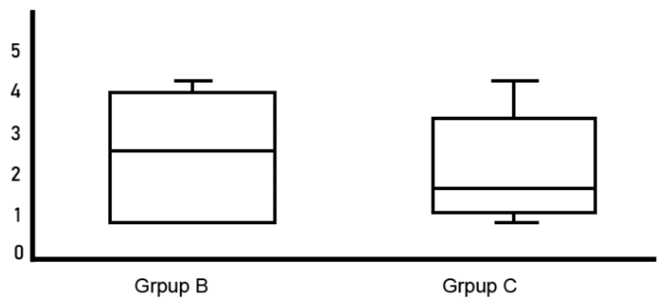
How pleasant the activity was (1 pleasant,5 unpleasant)			
Scale	Group A	Group B	Group C
1	21.05 %	10.34 %	-
2	42.11 %	13.79 %	35.71 %
3	15.79 %	10.34 %	32.14 %
4	21.05 %	13.79 %	14.29 %
5	-	51.72 %	17.86 %
How easy was the task to perform (1 easy, 5 difficult)			
1	22.22 %	9.38 %	13.04 %
2	44.44 %	6.25 %	8.70 %
3	33.33 %	9.38 %	78.26 %
4	-	12.50 %	-
5	-	62.50 %	-
How easy it was to learn the task (1 easy, 5 difficult)			
1	53.85 %	16.67 %	13.04 %
2	46.15 %	-	26.09 %
3	-	-	26.09 %
4	-	-	34.78 %
5	-	83.33 %	-

6.5.5. Accessibility Up/Down Button

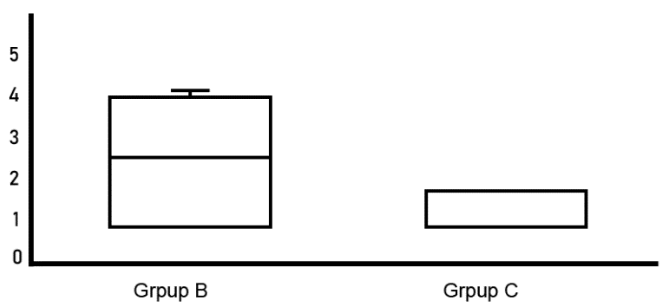
To ascertain the perception of participants of groups B and C participants on the accessibility feature for adjusting the height/elevation using the up/down buttons, we asked the same three questions as in section 6.5.2 (Figure 58a-c).



(a)



(b)



(c)

Figure 58.Distribution of responses

For groups B and C on accessibility buttons. (a) How pleasant, (b) How easy to perform, (c) How easy to learn

The results from the normality tests showed that the data is not normally distributed ($p < 0.05$), indicating that the responses have violated the assumptions for a parametric test. The Mann-Whitney U test was selected to analyze this data. This analysis helped to assess the mean rank differences.

From the output of the analysis (Table 22), it was recorded that for the question "How pleasant the activity was", "How easy it was to perform the task" and "How easy it was to learn the task", there is no significant difference in the mean rank for the distribution of responses ($p > 0.05$).

Table 22. Summary of the Responses by Groups B and C on the Aaccessibility Button

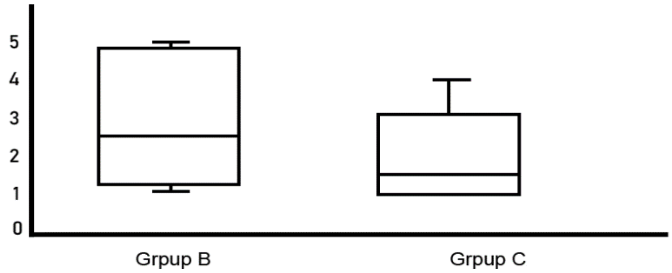
How pleasant the activity was (1 pleasant,5 unpleasant)		
Scale	Group B	Group C
1	10.34 %	27.78 %
2	13.79 %	33.33 %
3	10.34 %	16.67 %
4	13.79 %	22.22 %
5	51.72 %	-
How easy was the task to perform (1 easy, 5 difficult)		
1	13.79 %	11.54 %
2	6.90 %	23.08 %
3	-	-
4	27.59 %	46.15 %
5	51.72 %	19.23 %
How easy it was to learn the task (1 easy, 5 difficult)		
1	13.79 %	26.32 %
2	6.90 %	21.05 %
3	-	-
4	27.59 %	-
5	51.72 %	52.63

6.5.6. Accessibility Slider Feature

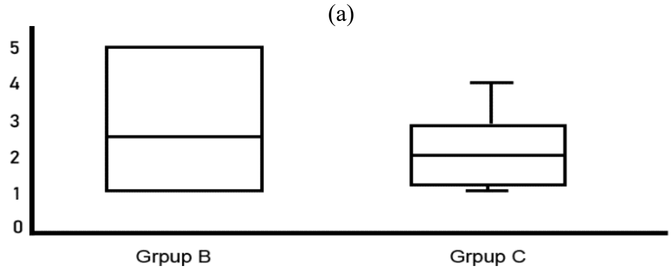
To ascertain the perception of participants of Group B and Group C participants on the accessibility feature for adjusting the height/elevation using the slider feature, we asked the same three questions as in section 6.5.2 (Figure 87a, 87b, 87c).

The results from the normality tests showed that the data is not normally distributed ($p < 0.05$), indicating that the responses have violated the assumptions for a parametric test. The Mann-Whitney U test was selected since there were only two groups presented. This analysis helped to assess the mean rank differences.

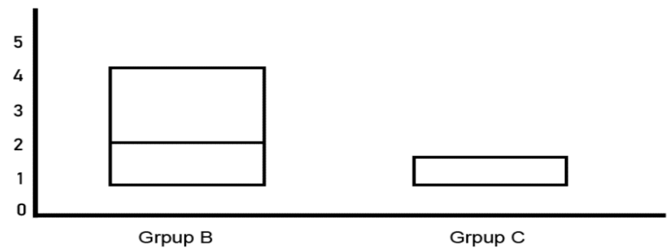
From the output of the analysis (Table 23), it was recorded that for the question "How pleasant the activity was", "How easy it was to perform the task" and "How easy it was to learn the task", there is no significant difference in the mean rank for the distribution of responses ($p > 0.05$).



(a)



(b)



(c)

Figure 59.Distribution of responses for groups B and C

For accessibility "Slider" feature. (a) How pleasant, (b) How easy to perform, (c) How easy to learn

Table 23. Summary of the Responses by gGroups B and C on the Accessibility Slider Feature

How pleasant the activity was (1 pleasant,5 unpleasant)		
Scale	Group B	Group C
1	10.34 %	25.00 %
2	13.79 %	10.00 %
3	10.34 %	45.00 %
4	13.79 %	20.00 %
5	51.72 %	-
How easy was the task to perform (1 easy, 5 difficult)		
1	13.79 %	14.29 %
2	6.90 %	38.10 %
3	10.34 %	28.57 %
4	-	19.05 %
5	68.97 %	-
How easy it was to learn the task (1 easy, 5 difficult)		
1	17.24 %	22.22 %
2	-	55.56 %
3	-	-
4	13.79 %	22.22 %
5	68.97 %	-

6.5.7. *In-App Data (Efficiency/ Task completion rates/ Accuracy)*

As explained in Chapter 3, Section 3.2.4, the in-app data-collecting system was used to track and record users' paths and interactions with the environment and virtual objects. This feature allowed us to measure users' performance, completion rate, and accuracy of each task. The interaction data on the device was sent to the researcher's email upon the completion of the VR experiment with a unique ID generated automatically by the app. Participants would use this ID to complete a post-experiment survey to associate their VR experiment with the online survey.

As seen in Table 24 below, it was noted that most of the participants completed or mostly completed the three VR activities with different levels of accuracy. The completion rate classified as "incomplete" indicates that none of the activities were completed. "50 % completed" shows that only the first two activities were completed and the last was either minimally done or not attempted. "Mostly completed" represents participants that completed all three (3) activities with a few skipped steps. The category "completed" indicates that all three activities were completed. The accuracy level of tasks was evaluated based on 3 levels: High (H), which means that the sequence of the task execution was done correctly, and the objects were placed in the right order and placement. Moderate (M) means that order and placement were mostly accurate with a few missed steps, and Low (L) means that the order and placements were not as accurate with several missed steps.

Each activity data was analyzed with a relevant method, including parametric and non-parametric Analysis of variance test (ANOVA) to investigate the differences in the groups.

Table 24. Summary of the Completion Rate/Accuracy

Completion rate/ Accuracy level							
Group	Incomplete	50% complete		Mostly complete		Complete	
A	-	20 %	M	40 %	M	40 %	H
B	-	10 %	M	20 %	H	70 %	M
C	10 %	10 %	L	20 %	M	60 %	M

The normality of the time spent on the tutorial activity was investigated using the Shapiro-Wilk test. The results from the test showed the data is not normally distributed. Thus, a non-parametric approach was adopted to assess the differences between the three groups concerning the time spent on the tutorial activity. The difference in the groups was analyzed using the Kruskal-Wallis test. For the tutorial activity, it was noted that there was no significant difference ($p > 0.05$)

in the distribution of the tutorial activity duration for the three groups of participants, indicating that time spent on the tutorial activity did not differ for the three groups.

The normality for activity 1 duration was assessed using the Shapiro-Wilk test. From the results of the test, it was noted that the data is normally distributed. Therefore, a parametric approach was adopted to assess the differences between the three groups for the time spent on activity 1. The difference in the groups was analyzed using ANOVA. For this activity, it was noted that there is no significant difference in the ($p > 0.05$) average time spent on activity 1 for the three groups of participants, an indication that time spent on Activity 1 did not differ for the three groups (Figure 60).

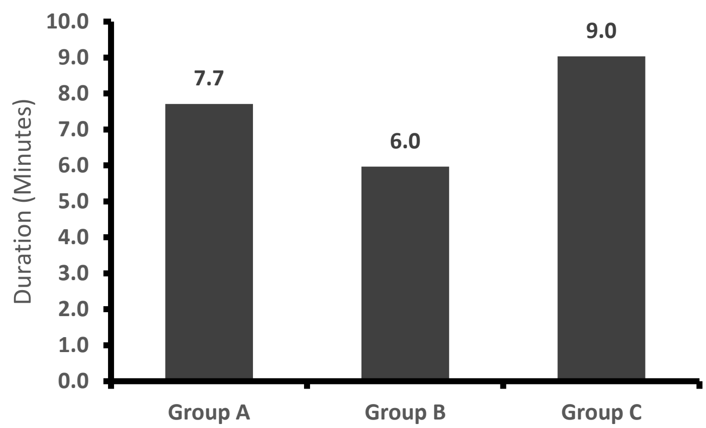


Figure 60.The average duration of activity 1 (in minutes) for the three groups

The normality of the time spent for activity 2 was evaluated using the Shapiro-Wilk test. The results from the test showed the data is not normally distributed. Thus, a non-parametric approach was adopted to assess the differences between the three groups, concerning the time spent on activity 2. The difference in the groups was analyzed using the Kruskal-Wallis, a non-parametric ANOVA test. For activity 2, it was noted that there is no significant difference ($p > 0.05$) in the distribution of the activity 2 duration for the three groups of participants, indicating that the time spent on activity 2 did not differ for the three groups.

By analyzing the activity differences within each group, we noticed that all three groups spent the most time on activity 1 and the least time on the tutorial activity.

For the group differences based on the total duration of the experiment, the Kruskal-Wallis test was selected to evaluate the distribution differences. It was noted that although group C spent a longer time completing all three activities, there is no significant difference ($p > 0.05$) in the distribution of total duration for the three groups (Figure 61).

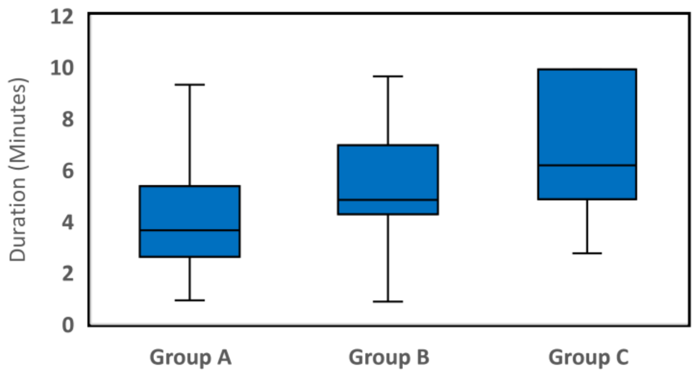


Figure 61.The duration distribution (in minutes) for the three activities

6.6. Discussion

6.6.1. Hypotheses

The results of this study helped us to answer two research questions as follows:

RQ5: Can software controls for vertical movement provide improved accessibility for wheelchair users with task efficiency and completion rate similar to non-wheelchair users?

The result of this study shows that 80% of wheelchair users could "complete" or "mostly complete" the tasks with moderate to a higher level of accuracy. This is comparable to 90% for the verifying group (non-wheelchair users who used accessibility features) and 80% for the control group (non-wheelchair users) not using any such features achieving a similar level of performance and accuracy confirming our H1 that accessibility features improved task completion rate and accuracy for wheelchair users.

The open question about participants' experiences in the VR lab revealed that it "took a while to learn" how to do things, then became easy once they learned how to use the controller. Some participants found it "very realistic; the size, shape, orientation, and placement of the objects around the lab". A participant commented that "as a first-time user", she was able to quickly learn hand motions and button combinations.

RQ6: Can the improved accessibility (if any) improve learning?

As shown in the result section, the question of "how easy it was to learn the task" indicates that although the verifying group and wheelchair users did not find it as easy to learn the activity (compared to the control group), both groups were able to complete the tasks with an almost similar level of accuracy and completion rate. They merely spent more time in VR to achieve similar outcomes. The average time spent in the VR for the control group was 16 minutes vs 17 minutes for verifying group and 21 minutes for wheelchair users.

Similarly, depending on the requirement to lower or raise the body to perform a task, 68% of wheelchair users found it moderately easy (rating 2-3 on the Likert scale) to perform the tasks that required lowering the body using the accessibility tool. However, only 21% of this group found it moderately easy to perform the tasks that required raising the body.

Several comments obtained through the open question indicated that participants believed that a VR lab and experience like this "can positively help in all areas of science and accessibility". They found the experience very "fascinating and fun". While some users found certain parts of the object manipulation tasks challenging, several participants agreed that it was a "fun and engaging" experience that helped them understand the topic of this chemistry experience better. From the data analysis, we can notice that for the accessibility feature, software control is appreciated more in group C and that the data is not significantly different from group B (verifying group) confirming our H2. Even though the sample size is small, we can still notice a definitive trend if more samples are taken.

The vertical movement with two methods (UI button and slider) was studied and tested with three groups of wheelchair users vs. non-wheelchair users. The higher rate of application for the Up/Down feature shows that users preferred this feature compared to the slider feature rejecting our H3 that the slider will be preferred compared to the UI button). It is worth mentioning that we have added the up/down control to the joystick button (hardware control) in the latest version and we are conducting further (unofficial) user testing (work in progress.

The overall data suggest that group B has more in common with group C (vs. group A). VR prior experience can be one of the reasons as most of the VR-experienced participants were in group A. Although it is not significant enough, perhaps collecting more samples would show a larger impact. Having more experienced VR users in group A was a coincidence. The lack of significance could be because there is not enough data to verify the claim.

6.6.2. Limitations

While many of the technical issues and bugs were fixed for this version of the study, updating VR components and plugins, headset firmware, and the Unity game engine remained a major issue that required code updates.

Due to the restrictions and limitations during the pandemic, it was challenging to recruit wheelchair users and other participants in general for conducting user studies. Two (2) participants in our group C, were simulated wheelchair users. The result from one simulated wheelchair user was excluded due to cybersickness and incomplete experiments.

Accessibility is a major and broad topic, and this study only started the investigation into the effect of VR on increasing accessibility. Further research is needed to consider different accessibility concerns and solutions.

Chapter 7. Research Reflections

In this chapter, we describe our proposed theoretical framework to use VR effectively in lab-based science education, addressing our seventh and last research question:

RQ7: What are the specific guidelines for accessibility and general guidelines for designing VR

We propose *ScienceVR* as a STEM learning framework that aims to provide an authentic and customizable experience with features that make it more accessible and inclusive to a wider range of audiences including wheelchair users. *ScienceVR* is designed to be authentic, customizable, learning-centered, and accessible. Before discussing this framework, we will review our findings and limitations to gain an overall insight into the use of VR in STEM education.

7.1. Summary of Findings

Through three phases of this study covering VR interactions, learning and accessibility we can summarize our findings as follows:

- The pilot study demonstrated the practicality and feasibility of running a full VR experience on a mobile headset. The experiment was lightweight enough that could be easily downloaded by participants and easily installed on their headsets. The in-app data collection systems and online survey tools and knowledge tests worked as expected. With some modifications and fixing UX issues, the pilot study package was used for the next large-scale study (i.e. study 2a).
- Study 2a revealed that while users found 2D and video content easier to use, VR performed better in terms of task completion and learning outcome. Immersive VR performed better with more accuracy compared to Desktop VR in task completion as well. The usability survey showed a high level of satisfaction with VR and participants found it to be fun, engaging, and game-like therefore engaging.
- Study 2b revealed that VR had a positive impact on the learning outcome as seen in the knowledge test score and participants found the experience very engaging and helpful to understand complex concepts of SPS hybridization. While there was no preference in the use of interaction techniques by the participants, we learned that the ray-casting technique is more efficient compared to the body/hand movement interaction technique.
- Study3 showed that VR can be made accessible for wheelchair users by adding features that enable them to overcome limitations.

As the focus was on vertical movement a height change feature with two methods was tested and the results revealed that a high percentage of wheelchair users could complete or mostly complete the tasks with moderate to a high level of accuracy comparable to non-wheelchair users.

7.2. Limitations

During these studies, we encountered design challenges, technical issues, and implementation problems. Most of the technical issues related to the process of animating objects, interactions with objects, texture, and lighting were solved by consulting with experts and solving them. Other challenges included finding eligible participants with VR headsets.

Considering the remote nature of these studies, the qualitative data in the form of observation was missed and we had very limited opportunities during the ease of COVID restrictions to observe users during the VR experiment which did not exceed three cases. Minor hardware glitches were mostly related to the battery usage of the hand-held controller and replacing them with new batteries. Also, technical limitations and device performance issues prevented us to ask every participant to video record the process on each device which would result in generating a very large file that would not be transferable. For this reason, we had to solely rely on in-app data to determine the user journey through the virtual experiment and a limited number of observations.

Although users were asked to connect the headset to a reliable Wi-Fi before the start of the experiment, there were a few times that the internet connection was not established properly resulting in the error in sending the data remotely to the researcher's email. However, the log data was saved on the headset and could be later retrieved for analysis. And lastly, the participants in these studies consisted of only university students in the Chemistry lab course which limits the generalizability of the results. Future research may replicate this study by recruiting participants of various age groups in other educational levels such as high school and other STEM fields.

While several research studies have been reviewed and referenced in this document, many remain experimental with limited scope. As a result, the lack of a cohesive design framework for the implementation of VR in STEM education is a noticeable need. By relying on the outcomes of the studies completed in different phases of our study, we propose the *ScienceVR* framework in the next section. Further research on different STEM topics, accessibility issues, interaction methods, and privacy concerns is necessary but beyond the scope of one thesis.

7.3. *ScienceVR* Framework

7.3.1. Basic Model

The path to the development of our proposed framework started by focusing on the specific challenges of the concepts, environment, content delivery, and training requirements for lab-based courses. The next step was to investigate VR affordances and the characteristics of an immersive VR solution and its challenges. In the final stage, we defined a set of criteria for the *ScienceVR* framework to be focused on learning, authentic, and accessible with an embedded feature for user data collection (Figure 62). This framework connects the human component (i.e., the learner), to the learning topics in each STEM domain via a VR headset (as the technology component of the framework). The proposed topics are curriculum-specific meaning that they are designed based on the content and materials students are expected to learn during the academic term.

In this thesis, we covered the field of Chemistry as an example domain for STEM education. However, the *ScienceVR* framework is characterized by its flexibility/ modularity that can expand and cover other domains such as Physics and Engineering. It can be a practical and affordable mode of content delivery that can be implemented locally (Standalone headset implementation) or be a cloud-based solution in the case of using WebXR technology[23].

This framework can include smart features such as Artificial Intelligence (AI) and Machine Learning (ML) systems based on the data generated by user interactions. The data collection system is an important part of the smart VR component to collect users' data and interactions which (in the future) can be ingested and analyzed by an intelligent system within the framework. This system can potentially make recommendations based on the learner's progress and achievement level. The highlighted areas in Figure 62 denote the path completed by our studies and have been tested during these studies. The assessment, recommendation, personal feedback, and content/experience improvement are examples of what can be done if we obtain learners' data. Collecting data from many users allows using AI methods to model them, find patterns, and then correlate the behavior patterns with results like academic success. While we see the value of data collection and that is why we created our app and used data collected in the framework, the actual data analysis with or without AI is out of the scope of this thesis.

---

[23] WebXR Device API is a Web application programming interface (API) that describes support for accessing augmented reality and virtual reality devices, such as the HTC Vive, Oculus Rift, Oculus Quest, Google Cardboard, HoloLens, Magic Leap or Open-Source Virtual Reality (OSVR), in a web browser. (WIKIPEDIA)

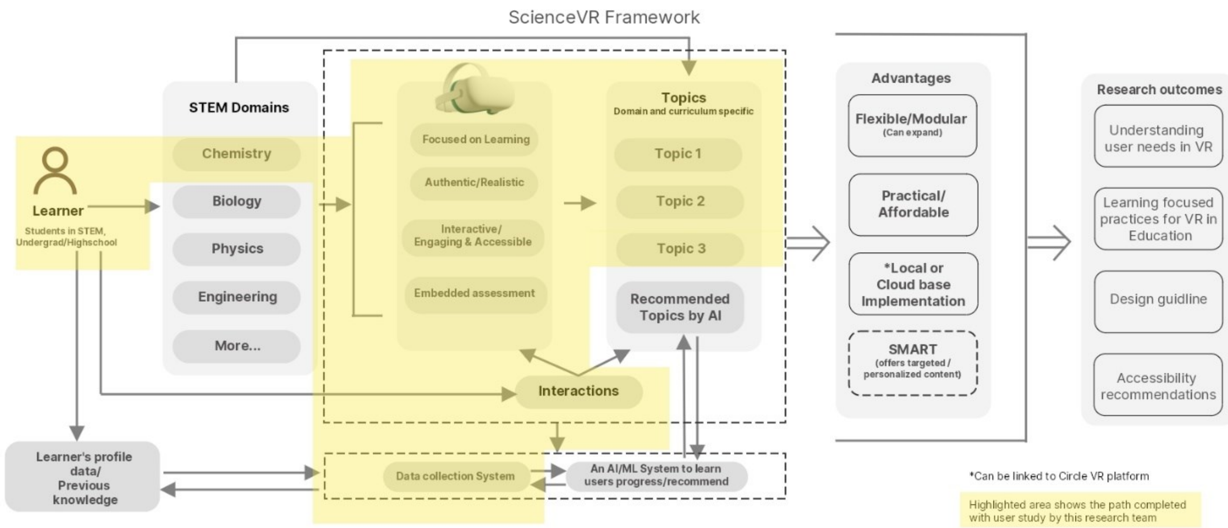


Figure 62.Full view of *ScienceVR* framework.

The in-app data collection system is embedded in this framework that does not rely on any third-party Application Programming Interface (APIs). This feature allowed us to track and record users' interactions to gain a full picture of their VR experience which can be used to provide reliable data for assessing learners' performance in the virtual space. This framework seems to be a promising and timely solution, especially during the current condition of the COVID-19 pandemic and the need for richer e-learning systems.

Building on the solid and established learning theories explained earlier, *ScienceVR* has the potential to go beyond applying mere basic interactions and realistic surroundings which are the most frequently used design elements in VR experimental studies. With *ScienceVR* we apply all relevant design elements that can create a completely immersive VR experience. The authenticity of the *ScienceVR* experience comes from the highly realistic, immersive, and interactive nature of the environment that allows learners to engage with the content as they would in a physical science lab. On the other hand, the virtual and component-based nature of *ScienceVR* makes it customizable to simulate real labs or even situations that are not possible to have in the physical world. *ScienceVR* is designed based on well-established learning theories and on seven VR design elements [10]: basic interaction, immediate feedback, virtual rewards, realistic surrounding, use of audio (feedback or narration), moving around, and knowledge test. The in-app data collection system is a part of *ScienceVR* that makes the data collection and metric-based learning assessment more feasible compared to relying on external Application Programming Interfaces (APIs).

Furthermore, this framework makes the VR environment more accessible to a wider range of audiences with mobility issues (i.e. wheelchair users).

As seen in Table 2 (STEM characteristics and VR features in Chapter 2), we can categorize STEM requirements into the following areas addressed by *ScienceVR* design:

- **Infrastructure and equipment:** *ScienceVR* is an inexpensive and feasible solution that can function as a complementary method providing somewhat close to the real experience needed for students in a science lab. It is configurable, customizable, and reusable and can be accessible making it a suitable alternative to physical labs.
- **Safety procedures:** We can simulate any dangerous scenarios in the *ScienceVR* framework to create skills and safety training procedures. The application of the headset and running any simulation is also safe (within the limits of VR safety regulations) as it does not require a powerful computer tethered to the headset hence reducing the risk of colliding with the environment (i.e., indoor furniture, etc.)
- **Content/subject matter:** *ScienceVR* is designed based on learning theories and the direct involvement of domain experts. It is an expandable platform that can potentially include any course-based content currently being taught by the instructors.
- **Active learning:** *ScienceVR* is a fully immersive platform that allows users to use their natural body movement, and hand gestures, walk or move around using teleportation and be able to interact with the environment.
- **Assessment:** A system of an in-app quiz and data collection provides a way for educators to collect, analyze and assess the learning outcome from the *ScienceVR* framework making it a unique platform that does not rely on the third-party/external APIs to collect user's data.
- **Accessibility:** With features such as users' height adjustment, alternative interaction methods, and visual enhancement for people with Colour Vision Deficiency, *ScienceVR* will be accessible to a wider range of audiences who are currently excluded or partially excluded to enjoy a fully immersive VR experience.

Details on the educational theories and how it was used in building this framework are provided in the design and are shown in Table 25. Although the *ScienceVR* prototype is developed with undergraduate students in Chemistry or any other Science or Engineering program in mind, it can be used by learners in general. Future development plans can allow anyone with access to a VR headset or desktop computer to use it without the need for previous experience.

Course instructor who manages the selection of content and potentially other developers who would be able to add new content are the secondary users of this framework.

7.3.2. ScienceVR Design Principles

To help guide the accessibility considerations we use Universal Design[24] and Universal Design for Learning (UDL) guidelines and related principles[87], focusing on content and interaction as key areas[86]. The UDL Guidelines are tools used in the implementation of the Universal Design for Learning, which is a framework to improve and optimize teaching and learning for all people based on scientific insights into how humans learn [121]. Our proposed framework is built on the STEM education characteristics and affordances of VR shown in table 3 we are using Chemistry lab safety procedures as a suitable example for building a prototype and conducting user testing. In phase one we deal with authentic simulation as mentioned in the STEM education characteristics that require physical laboratory space and equipment; learners need to conduct scientific experiments to gain required skills and the fact that safety procedure is a major component in working in science laboratories.

ScienceVR design features are considered to address our basic goals to be authentic, customizable, learner-centered, and accessible. Each feature corresponds to one or more applicable learning theories and design goals. From the technical point of view, it is a multiplatform framework that supports HMDs and desktop VR providing a suitable environment for learners to experiment. It is more accessible and can be customized based on learning goals. The realism and assessment features are also included to achieve authenticity and learning-centered design goals. All relevant learning theories described in the previous section and how they are mapped to the design goals are highlighted in Table 25.

[24] http://universaldesign.ie/What-is-Universal-Design/The-7-Principles/

Table 25. Features of the *ScienceVR* Framework

Feature	Motivation	Learning theory	Design goal
Multi-platform with HMD VR support	Immersive VR is engaging spaces and scenarios Should be accessible	Constructivism, Active learning	Accessibility, Customization
Multiple-platform with desktop VR support	More accessible Familiar platform, no need for training new controllers Limited immersion	Constructivism, Active learning, accessibility	Accessibility, Customization
Virtual Environment	Feasible to create Authentic immersive experience	Constructivism, Active learning	Authenticity, Learning-centred
The in-App data collection system	Assist in learning assessment	N.A.	Learning-centred
Interaction			
Locomotion (Travel, Teleport, Walk)	Users have options to select any mode to move around the VE Teleport navigation can reduce motion sickness	Experientialism	Accessibility, (Usability)
Hover (Ray-cast), selection, grasp	Simple interactions lower cognitive load	Cognitivism, Experientialism	Learning-centred, Accessibility,
Learning			
Levels / staged progress	Starting with lower fidelity to direct the learner's attention to the most important elements Help to build skills Reduce training time and cognitive load	Scaffolding, Cognitivism	Learning-centred, Accessibility,
Interactive text/help tips	Providing information when needed Reduce cognitive load	Segmentation (Chunking), Cognitivism	Learning-centred, Accessibility,
Sound, haptic feedback, visual highlights quiz, and virtual rewards	Provides immediate feedback Reduce cognitive load	Guided learning, Instructional Strategies, Signaling, Gamification	Learning-centred, Accessibility,
Future			
Adding intelligence/AI for customizable/ personalized learning VR			

The following sections provide specific recommendations and requirements for VR design and development in STEM education, accessibility in VR, and general VR development. The requirement levels for a VR implementation are defined in Table 26.

Table 26. Requirement Levels

Term	Description
Must have (MH)	This requirement is mandatory, an essential part of the development
Nice to have (NH)	If there is a time and option to add, this can be included
Future plan (FP)	Time and budget constraint do not allow at the moment but can be considered in the future when the constraints are removed

7.4. Recommendations for *ScienceVR* Design and Development

Many of the standard guidelines in the design and development of VR experiences relate to the technical aspects of the systems to enhance performance (see details in section 7.7). While it is very important to improve the performance component in VR (to avoid issues such as latency for example), more is needed to be done to expand standards and guidelines for designers and developers to cover other aspects such as user experience (UX) and accessibility. In essence, the focus should be on creating a comfortable and safe VR experience, and maintaining a high frame rate is a part of that effort. The reduced frame rate can cause latency and consequently induce cybersickness.

Some of the general tips provided for Performance Targets[25]:

- **Frame per second (FPS):** Media applications are allowed to target 60 Hz FPS while interactive applications must target a minimum of 72 FPS. The use of profiling tools for CPU/GPU usage is recommended to check rendering statistics and to pinpoint and address issues causing dropped frame rates.
- **Draw calls:** Every visible object in a virtual scene is sent to the graphic card to be drawn. This can take lots of processing power if there are many objects in a scene. Therefore, the request to draw them on GPU will be very high which can cause latency. Merging objects and using baked lights and textures for a group of 3d objects can reduce the number of draw calls thus improving performance.
- **Mesh economy:** It relates to the number of triangles count for 3D objects (Figure 63). Similar to draw calls, an increased number of triangles can negatively impact the performance of a VR experience. The use of a mesh optimizer and other techniques can reduce the number of triangles for better performance.

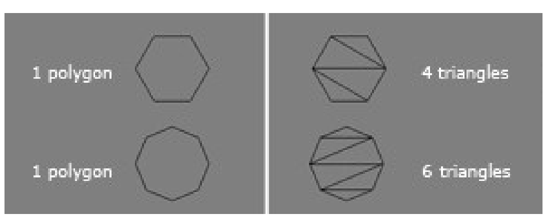


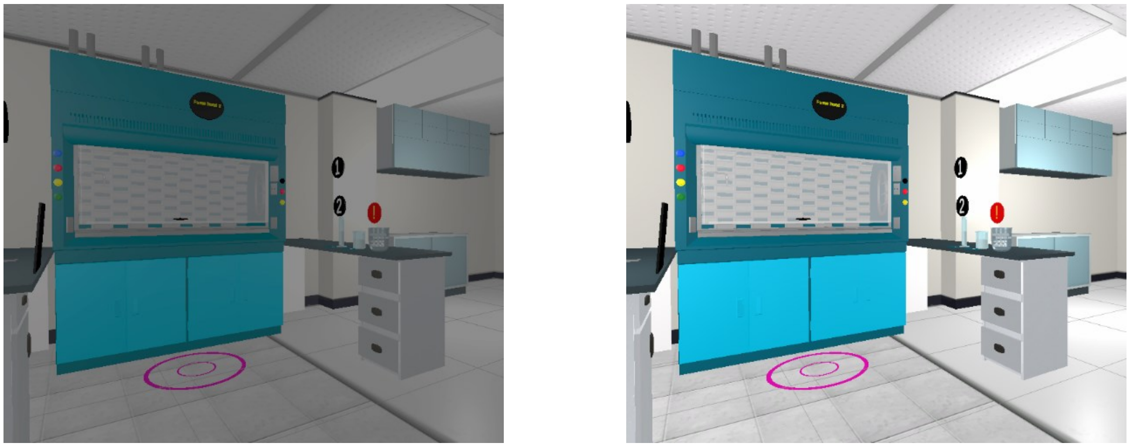
Figure 63.3D Objects are created by the triangle shape

Credit: http://wiki.polycount.com/w/images/f/f6/Triangles.jpg

[25] As we have used oculus quest headset in our studies, we only included the recommendations for this device. This recommendation would work with other headsets with similar specifications.

Some of the techniques that can help optimize any VR implementation include:

- **Light baking** to reduce computational efforts to render a scene (Figure 64)



(a) (b)

Figure 64.Baking the light (a) before (b) after

- **Occlusion culling:** To be used to avoid rendering objects that are out of the view of the camera, therefore, reducing the computational expense of the Central Processing Unit /Graphics Processing Unit (CPU/GPU). This feature can be activated for any object in the scene (Figure 65).

Figure 65.Occlusion culling feature

(To prevent rendering objects that are not in the view)

- **Static batching:** By setting all stationary objects in the scene static, GPU/ CPU performance will improve. This can be done by tagging any object as static. It will also reduce draw calls and other calculations on the objects.
- **Quality settings:** Setting the quality of objects being rendered to "Fast", "Simple", "Good", and texture quality to "Full" or "Half" resolution can also improve performance.
- **Rendering pass type:** Unity supports Multi-pass, Single-pass and Single-pass instanced to render the scene. Multi-pass renders most objects twice to generate more accurate lighting at a cost of higher processing power. Single-pass renders two textures into one big texture and has a better performance however it is not available on all devices. Single-pass instanced has added reduced GPU overhead compared to Single-pass.

We have used single-pass renders in *ScienceVR* implementation with good results and recommend it to be tested in most of similar scenarios.

Other potential issues to be considered during the design and development:

- **Decide on the appropriate render pipeline** that meets the requirements (universal render pipeline vs high definition detailed below) in case the Unity game engine is used for the development. As defined by Unity, a Universal render pipeline (URP) is a prebuilt scriptable render pipeline that provides better performance for creating optimized graphics suited for a range of platforms including mobile, and VR headsets.

  High-Definition Render Pipeline (HDRP) on the other hand is a high-fidelity programable pipeline better suited for a modern platform that is compatible with procedural shaders. The main difference between URP and HDRP is the performance as HDRP uses physically-based lighting that requires more processing powers needed in modern games, technical demos, and high-quality animation.

- **Use of third-party plugins, shaders, and data collection systems**

  For creating more complex shaders such as realistic lighting and reflections, the use of procedural shaders by third-party plugins (such as shader graphs[26]) will help to create more appealing and aesthetically pleasing content. However, some of those shaders may not be

[26] https://unity.com/features/shader-graph

compatible with the type of render pipeline discussed in the previous paragraph. This should be considered at the design stage of the VR experience.

Similarly, some of the third-party data collection systems / APIs to track and record user data (such as Google Firebase[27]) may be helpful but can be costly as well.

- **Multiplayer capability**

  Multiplayer capability can take a VR experience to the next level in terms of interaction and social aspects of learning. This feature and the possibility of adding it to a VR development project should be considered during the design stage.

7.5. Recommendations for Accessibility in VR

As the use of VR technology grows the importance of making this technology accessible to more people will increase. VR hardware manufacturers and software companies[28] are beginning to introduce measures and recommendations to encourage developers of VR content to consider ways for making their software more accessible. Although these attempts are a move in the right direction, it is not nearly enough to address the need for accessibility in VR implementations.

One of the major concerns is the accessibility consideration often lags the development cycle and the other is a patchwork that is added to the product in later stages of production.

Some of the aforementioned recommendations include the use of:

- Visual indicators such as highlights for interactable objects
- Spatialized audio
- Haptic feedback
- Comfort and locomotion settings
- In-game subtitle
- Color vision deficiency features
- Height adjustment features for wheelchair users

An example of each recommendation is shown in Table 27.

- **Visual indicators** such as highlights for interactable objects.

[27] https://firebase.google.com/docs/storage/unity/start

[28] As of November 2020, Virtual Reality Checks (VRCs) was announced by Meta (Oculus/Facebook) to make VR applications more accessible and inclusive for "all people". VRCs is a set of technical recommendations to assist developers create more accessible VR experiences.

- **Spatial audio** in virtual reality is defined as the manipulation of audio signals to mimic acoustic behavior in the real world. An accurate audio representation of a virtual world adds up to a compelling and immersive experience.
- **Haptic feedback:** Use of vibration patterns and waveform to convey information of collision/contact with virtual objects in VR.
- **Comfort and locomotion settings:** It includes a set of features that allows users to adjust their comfort level for turning, movement, posture, audio, difficulty, etc.
- **In-game subtitle:** As an accessibility feature for dialogue and interfaces. This feature is gaining more interaction to help users with hearing impairments participate in VR forums.
- **Accessibility feature for Color Vision Deficiency (Color blinds).** This feature can be added to a VR project in Unity using a color-blind package.
- **Adjustable player height** to allow enhanced accessibility to objects that are beyond the reach of wheelchair users and other users with short heights.
- **Proper font size, color, placement for text, and UI elements** to allow maximum legibility in VR.

The text should follow the standard and recommended font size. Microsoft recommends for the near interaction at 0.45 m (45 cm), the minimum legible font's viewing angle and the height are 0.4°-0.5° / 3.14–3.9 mm. It's about 9-12 pt. with the scaling factor introduced in Text in Unity[29].

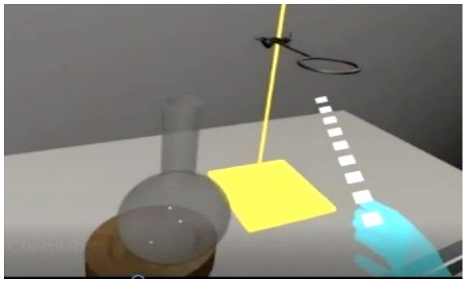
The minimum readable size for the font in VR is 1.330 or a height of 2.32 cm at a distance of one meter. The recommended size of the font is 3.450 or a height of 6.04 cm over a distance of one meter[30].

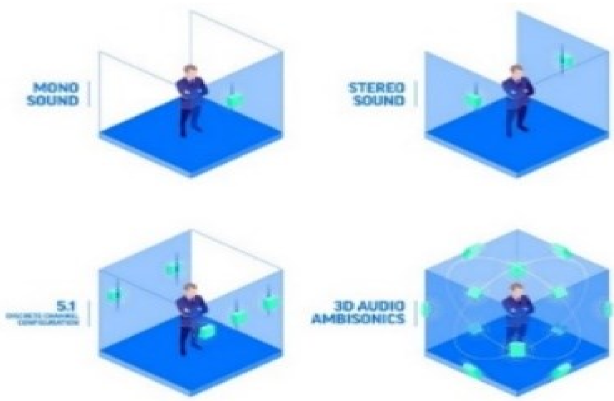
While the above-mentioned guidelines for font size and distance may work in most cases, they might not work in certain cases where users need to increase the text legibility based on their eyesight if needed. Therefore, we propose adding a slider that the user can use to adjust the size of each text panel.

[29] https://docs.microsoft.com/en-us/windows/mixed-reality/design/typography

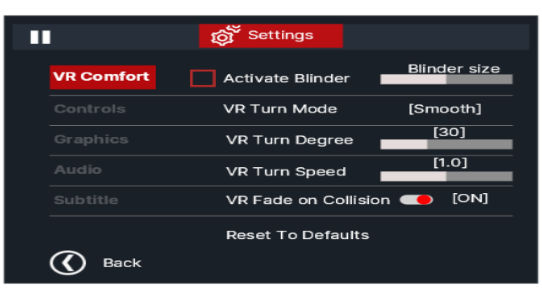
[30] http://vovakurbatov.com/articles/10-rules-of-using-fonts-in-virtual-reality#:~:text=1)%20Minimum%20readable%20size%20for,the%20distance%20of%20one%20meter.

Table 27. Accessibility Recommendations for VR

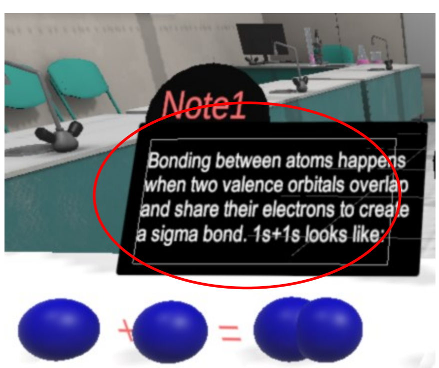
<table>
<tr><th>Example</th><th>Requirement level</th></tr>
<tr><td>


Highlighted objects
Pointing a hand-held controller towards the stand shows a yellow highlight, indicating that it is interactable
</td><td>Must have
(Functions as visual feedback)</td></tr>
<tr><td>


Different audio configurations including Spatial audio
Credit: https://tricktheear.eu/spatial-web-audio/
https://creator.oculus.com/learn/spatial-audio/
</td><td>Must have
(If not a full 3D version, some level of sound implementation in VR is required to enhance the immersive experience)</td></tr>
<tr><td>


Haptic feedback feature
Vibration can be activated on a hand-held controller when interacting with 3D virtual objects
</td><td>Must have
(to provide an extra layer of feedback to users when they interact with objects)</td></tr>
<tr><td>


Proposed VR comfort setting for a VR implementation
</td><td>Must have
(see further details in the next section)</td></tr>
</table>

 Use of subtitles in a VR online conference	Future plan (Can be added in future implementations)
 Use of color-blind unity package in VR Credit:https://www.alanzucconi.com/2015/12/16/color-blindness/	Must have (For people with CVD), making sure that no essential information in VR is conveyed by color alone.
 Accessibility features to change user height	Must have (For wheelchair users, those who may temporarily use a wheelchair, or people with shorter height
 Text and other UI elements The text should be the right size, with high contrast for legibility. It should be moved to the top layer when invoked to avoid being blocked by another object in the scene.	Must have The text content and UI elements should always be on the front layer to avoid being obscured by other 3D objects and follow the color and size guidelines for maximum legibility.

7.5.1. Cybersickness in VR and Design Considerations

Different studies suggest that up to fifty percent (50%) of VR users may feel some degree of cybersickness with symptoms such as nausea, headache vertigo, etc. In our three studies during this research, while testing over two hundred (200) users, we noticed that 10% of our users expressed some level of cybersickness to VR. Three percent (3%) of them had a rather severe reaction to the extent of not being to complete the experiment and did quit.

Many factors that can cause cybersickness in VR include sensory conflict, vection, flicker, and postural instability. Sensory conflict happens when information from the vestibular system does not match the information coming from our vision system. Vection is the illusion of self-motion induced by an optical flow pattern for example watching a moving train "creates the illusion that one's own stationery train is moving" [122].

Digital display flickers can cause eye strain and fatigue. Postural instability on the other hand involves not having effective strategies to maintain a stable posture. An example of this is a passenger sitting in a moving car. We noticed that using highly reflective materials for metal and glassware in the lab can create undesirable flickers. Using transparency and baked texture and lighting this problem can be mitigated without losing authenticity and many of the visual effects required to depict glass and metallic objects.

As illustrated in Figure 66, paying more attention to the comfort and locomotion settings in VR is very important with the goal to reduce the effect of cybersickness.

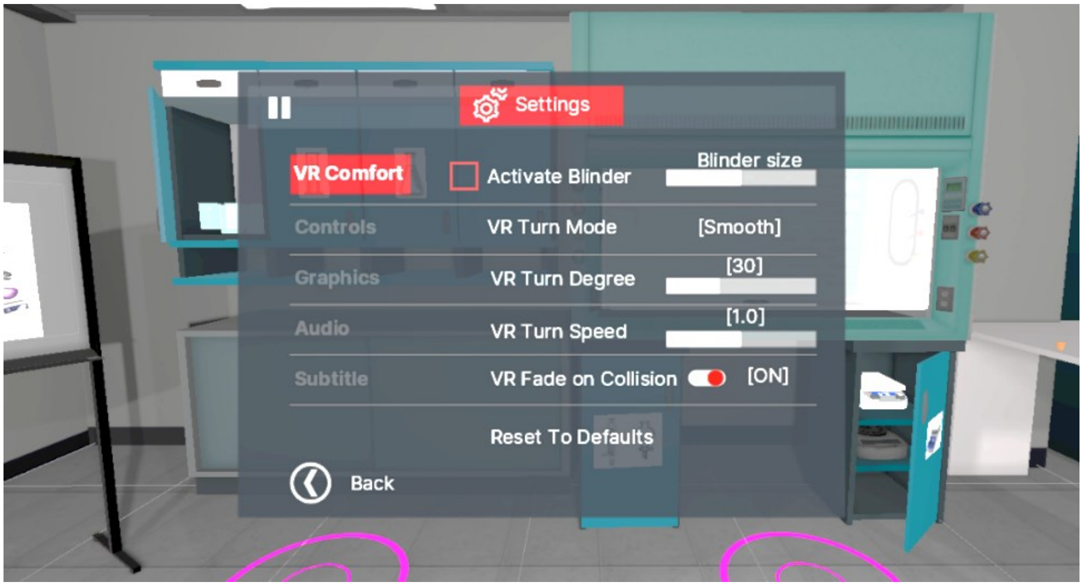
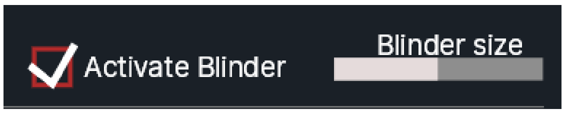


Figure 66.Proposed comfort settings for *ScienceVR*

7.5.2. VR Comfort Settings

VR comfort settings include a feature to activate a “Blinder”, which is a way to reduce vection by reducing the field of view (Figure 67). The size of this blinder can also be adjusted by using a slider seen on the UI screen.



(a) (b)

Figure 67.Blinder controller

(a) Moving the slider will increase or decrease the field of view (b) reduce vection

VR Turn Mode, Turn Degree, and Turn Speed are other features added to provide different methods of turning viewpoints with controllable speed for a smoother transition. Discrete rotation (known as rotation snapping) eliminates most of the optical flow of a continuous rotation to reduce cybersickness.

7.6. General Standards for VR Development

To develop VR experiences, there are certain level of general standards for software development in place (by the industry) that needs to be followed if the VR application is going to be available on a certain platform. For example, companies such as Meta (former Oculus/Facebook) define a set of technical requirements that all apps on the Oculus platform must meet to be published (detailed in the next sections).

The general standards we recommend for any VR undertaking that follows standard software development projects [122] include:

- A clear and concise description of the purpose
- Simplicity (of use and interactions)
- Reliability
- Efficiency

- Low cost
- Follow the established standards
- Accurate user documentation and technical documentation

All production and development levels must have a designated owner responsible for the tasks and processes critical to the project.

There must be a system of version control and separate production, development, and test environment to ensure security and backup versions of the production work as detailed in the next section. Software Development Lifecycle (SDLC) which is a systematic approach to creating software, can also apply to any VR implementation as described below:

- **Research to define requirements /features:** In this phase, all the requirements including project assessment, data storage, and retention requirements, legislative, security, privacy, access requirements, confidential data, access rights, and payment requirements (if this product is to be commercially released), login, reporting and training requirements, etc. should be defined and discussed for further analysis.
- **Analysis of users' needs, business goals, and market analysis:** Based on the research in step 1, data analysis is to be done to understand users' needs, business needs, and market conditions. In this phase, we should consider if all the requirements defined in step 1 can be implemented and tested or not. With the collaboration of domain experts in the related field of science, research should be conducted to understand the content design requirements. In this phase, functionality and technical specification including performance and design characteristics, environmental conditions under which the application is to perform, connection with external components/services, testing requirements, privacy and security specifications, user acceptance and implementation requirements, and the methods for evaluating users' satisfaction/efficiency of the system should be established and documented. Operational requirements such as data storage needs and potential growth, technology for user access remotely (e.g., VPN), and maintenance requirements also should be decided in this phase.
- **Design /Production (assets and coding specifications):** Designing the activities and creating graphic and 3D assets are done during this phase. Identifying activities, tasks, procedures, functions, control logics for each task, database and access requirements,

performance requirements, input, and output connections, and, error/exception handling are also to be completed and documented in this phase.

- **Development:** Coding to add interactivity, UIs, animations, etc. are done during this phase. Depending on the technology selected for the development, a proper technology can be used for coding defined in the requirement section. In hour case C# (Unity3D game engine) was suitable for local implementation and "A Frame/Three.js" seem good options for web implementation of VR. The codes should be written in a way that is understandable, easy to change (with a low risk of breaking the existing feature), and easy to troubleshoot.
- **Testing:** User testing debugging, and fixes happen during this phase. Each new change made to the product needs testing and documentation. Some features might require to be tested for accessibility compliance with accessibility-related organizations. Functionality and features are to be tested and the participants in each test should be identified and pass/fail criteria established before testing.
- **Implementation:** Deploying/releasing the product occurs based on a defined communication plan, to stakeholders, and the product is released on a scheduled date/time. The communication plan should include tasks required and team members involved to get the implementation phase completed. Risks and contingency plan associated with each risk is to be included. Any system configuration, support, training, and validation process should be detailed and explained clearly in the implementation plan.

Maintenance: To fix, change or enhance the product, if there is an agreement to continue support and maintain the product should be a part of the agreed maintenance plan.

7.7. Example Application of *ScienceVR* Framework

To apply lessons learned through this research and the guidelines proposed for our proposed VR framework, a collaborative effort is underway with Dr. Jim Davies and his student, Ludivine Blais, from the Department of Cognitive Science, and a fellow IT Ph.D. student, Anthony Scavarelli, who is studying VR in social spaces [20]. This effort is focused on developing BrainVR, a series of VR modules to illustrate the functions of various parts of the brain, The platform used in the BrainVR study is built using WebXR technology and not Unity (used in the previously described studies). But the principles and lessons learned from *ScienceVR* are applied in the BrainVR experiment.

BrainVR is primarily developed by our collaborators. The author of this thesis is acting as an advisor for VR design therefore the BrainVR experience is discussed here briefly and as an example only. The study is ongoing with preliminary results so far.

7.7.1. Overview

BrainVR is a new experiment using VR in a cognitive science concept called Memory Palace. Memory palaces have been a common learning memorization tool as far back as the ancient and Roman rhetorical treatises [123]. Often referred to as a method of loci, memory journeys, or mind palaces, memory palaces are a technique that takes advantage of the brain's natural impressive ability to use spatial information to recall a vast amount of novel and complex information [124]. In a memory palace, a list of things that are difficult to remember is encoded as images that are easy to remember. These images are placed in locations in the memory palace. The learner memorizes the palace and can walk through it in their imagination, decoding the mnemonics into the target information [125]. A mnemonic is an abbreviation of a more complex idea through symbols, letters, or images [126]. An example would be the mnemonic ROY-G-BIV to remember the colors of the rainbow in elementary school.

Memory palaces are typically modeled on locations well-known to the learner. Someone might make 30 canonical locations in the imagined version of their childhood home. Then, when they populate the locations, mnemonics are created and placed, in order, in these locations. It takes effort to create the palace locations, remember the correct order, create effective mnemonics, and rehearse the palace so that it is remembered when needed. This effort can make this memory tool less viable for many people. It could be that providing a palace for learners will reduce this burden.

While some (mostly older) literature uses the term VR only for fully immersive devices, the common trend is to use the terms Desktop VR (DVR) and Immersive VR (IVR) or Head Mounted Displays (HMD) VR [127]. These all refer to a VR experience accessed through different methods. Although DVR is less immersive, it is cost-effective and more accessible.

Related work on the use of virtual reality for memory retention has demonstrated that full immersion is significantly better than desktop [128]. Even abstract information, when experienced through a virtual lens, aids in the recall and understanding of the material. This is because immersion leads to more interaction with the data [129]. It is plausible that the learning of a memory palace could be enhanced if it is done through VR. This visually engaging module depicts the brain as a series of floating islands at different elevations. This enables islands to represent the brain's gross anatomy. The island positioning mirrors the actual relative placement of the parts in real brains. Each island has mnemonic scenes placed on it to represent key areas within those main brain parts. Each scene has two image elements: the foreground represents the function, and the background represents the name of the part.

7.7.2. BrainVR Experiment and Initial Results

To conduct a user study, we created a memory palace to teach students about brain regions for the first time including the name, function, and location of each brain part. Two experiments, Experiment 1 (N=15) and Experiment 2 (N=13) compared this brain VR learning tool to other traditional studying methods to assess its empirical value. Experiment 1 had three conditions: study as you wish (n=7), a study using text descriptions of the memory palace (n=5), and a study using images of the memory palace (n=3). Experiment 2 had two conditions: a study using a slideshow of the memory palace condition (n=6) and studying using a desktop VR version of the memory palace (DVR) (n=6). The headset VR condition (n=1) will be included in Experiment 2 in a subsequent iteration. Memory assessment was measured through a follow-up interview call 2 weeks later. We found no significance in either Experiment 1 or Experiment 2 between the type of studying method and interview participant scores.

General trends of Experiment 1 and Experiment 2 indicate that as memory palace elements were introduced, participant interview scores decreased. The current mean interview scores from highest to lowest are as follows: studying as you wish, using text descriptions of the memory palace, using images of the memory palace, using a slideshow of the memory palace, and a desktop virtual reality of the memory palace. Data collection will continue during the summer. Figures 68 and 69 show the scenes from the user's point of view in VR.

As this study will continue and more users will be tested in IVR and DVR, we will gain a better sense of the differences between using VR in this context.

Most of the recommendations made for the design of VR experiences (described in Section 7.5) are applied in BrainVR. These include the use of visual indicators for selected and interactable objects, use of audio (narration), haptic feedback, adding comfort setting (to minimize cybersickness), as well as several accessibility features including subtitle, text size change feature, color change settings (for color vision deficiency) and height adjustment feature for wheelchair users.

Figure 68.Study using DVR and IVR cerebellum stimuli

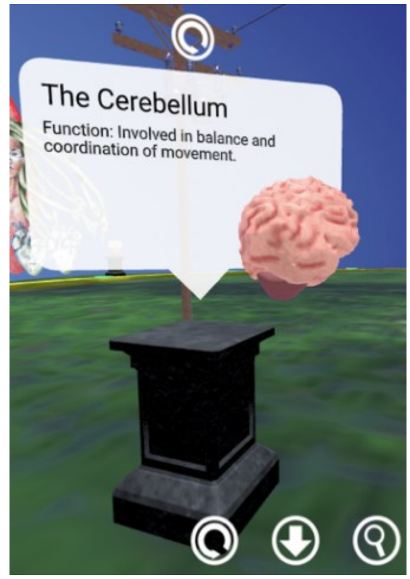


Figure 69.Study using DVR and IVR pedestal description

However, since the Web-based VR framework was used to develop BrainVR, the main technical principles in the development are web standards specifically the WebXR framework therefore some of the Unity game engine specifics such as the use of third-party shaders or occlusion culling feature might not be applicable in this case.

Regarding the application of learning theories in BrainVR, we made suggestions to improve the information presented to users that include:

- a tutorial section so that users learn how to use controllers (scaffolding)
- presenting information in small portions and when necessary (chunking and cognitive load)
- adding highlights to active/ interactable objects when hovered (signaling)

For improving user experience and accessibility, the following items are considered:

- adding text size change feature,
- higher contrast between written text and the background color,
- making sure that text always is in the users' field of view (on the top layer) so that the information is not obscured by other 3D objects in the scene (Figure 69),
- adding the height change feature for wheelchair users.

Chapter 8. Conclusion

To conclude this thesis, we summarize our experience in the design and development of a VR framework that resulted from the application of lab-based VR in three completed user studies. Through these studies, we gained new insights into the effect of VR on science education which led to defining a set of guidelines that should help future VR content designers and developers to build relevant, meaningful experiences for education.

By reviewing the literature on the use of VR in education, we identified several theories that could be relevant to VR-based STEM education. Next, we designed and developed a base VR system (in chemistry education as an example) for a pilot study to evaluate the feasibility and practicality of VR implementation on a mobile headset while adding a dedicated data collection system for online user testing. In three subsequent user studies, we explored and evaluated the effect of immersive VR on interaction, learning, and accessibility compared to 2D/Text-based and desktop VR.

To inform the design and development of a VR application in STEM education (for future VR designers and developers), we proposed a framework named *ScienceVR*. We then suggested a set of guidelines based on universal design principles, general software development, and accessibility considerations. An in-app data collection system and potential AI-based functionality of the framework will be essential components for creating a complete VR system with assessment, feedback, and recommendation features useful in educational VR applications.

Throughout our studies, we actively involved course instructors to help create relevant experiments. Their involvement will continue with the aim to obtain data and feedback from them that can be the subject of a new study recording their experience using VR as a teaching tool.

The results of our studies produced information on the feasibility and effectiveness of implementing VR to increase students learning outcomes and accessibility in science lab-based education. However, to achieve this goal many design and development considerations should be taken into the account. Through different phases of our study, we addressed some aspects of a useful VR implementation, more research needs to be done to expand and improve this framework.

Different STEM topics, accessibility issues, interaction methods, and also privacy concerns are among possible directions for future research to address the limitations of our studies.

Concerning the outline of the *ScienceVR* framework, we particularly envision two specific investigations:

1) to focus on improving the data collection system (Figure 70) to connect it to an external database with an admin portal (similar to Google Analytics). The collected data can then provide a visual analysis of users' interactions which can be used to provide an assessment/feedback loop on learners' academic performance.

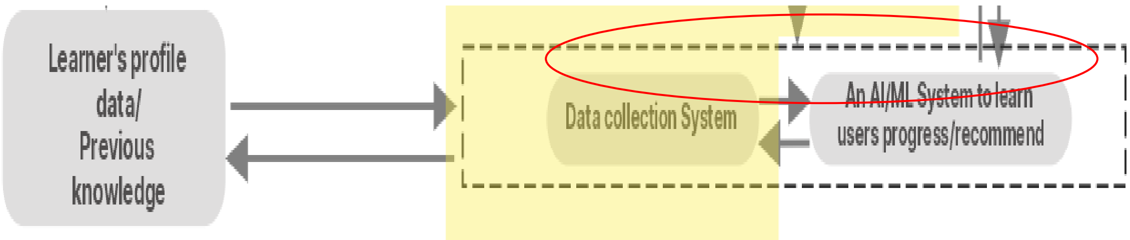


Figure 70.Future research

To focus on the data collection system and AI features of the *ScienceVR* framework

Considering this direction moving *ScienceVR* to a web-based VR platform such as WebXR (A-Frame[31] and Three.js[32]) may be a more feasible solution.

2) to add multiplayer functionality to the framework extending its reach and efficacy, considering the importance, need, and value of a collaborative experience in education [13]. An experimental effort by the researcher of this thesis is underway to test multiplayer functionality and the initial result is promising.

As the use of wireless VR headsets grows[33] new opportunities will be presented to conduct further studies using this technology.

Given the chance to redo our studies, there are a few things we could do differently. Experimenting with WebXR and web implementation (vs the current self-contained implementation) and adding multiplayer functionality from the start would be among those. Adding a real physical lab user testing to compare with virtual simulation would also help to gain

---

[31] https://aframe.io
[32] https://threejs.org
[33] https://techcrunch.com/2016/10/11/escaping-the-trough-of-disillusionment-for-virtual-and-augmented-reality

deeper knowledge on the efficacy of VR implementation. For a long time, VR was seen as a passive medium. Currently, we are seeing a shift towards what it can do differently and do it well, and that is, as Denny Unger[34] from oculus describes "physically suspending disbelief". No matter if the user is in a simple VR game or a complex flight simulation like Ace Compat (Bandi Namco), the total immersive aspect of the experience is something users can never get from a 2D screen experience. It may well be the right time for content designers, developers, publishers, educators, and educational institutions to give much-needed attention, investment, and effort to this technology not to replace the current educational content but to supplement it with VR.

[34] https://www.youtube.com/watch?v=mmv2uky4GhA

References

[1] A. Dünser, K. Steinbügl, H. Kaufmann, and J. Glück, “Virtual and augmented reality as spatial ability training tools,” in *Proceedings of the 7th ACM SIGCHI New Zealand chapter’s international conference on Computer-human interaction: design centered HCI*, 2006, pp. 125–132.

[2] L. W. Bailey, *Educational Technology and the New World of Persistent Learning*. IGI Global, 2019.

[3] K.M.Torres, A.Statti, “A Global Perspective of Classroom Technology Integration and Use, Educational Technology and the New World of Persistent Learning”, IGI, pp93-113. 2019.

[4] A. Y. Kolb and D. A. Kolb, “Experiential learning theory,” *Encyclopedia of the Sciences of Learning*, vol. 978, no. 1, pp. 1215–1219, 2012.

[5] C. Anthes, R. J. García-Hernández, M. Wiedemann, and D. Kranzlmüller, “State of the art of virtual reality technology,” in *2016 IEEE Aerospace Conference*, 2016, pp. 1–19. doi: 10.1109/AERO.2016.7500674.

[6] F. Buttussi and L. Chittaro, “Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario,” *IEEE Trans Vis Comput Graph*, vol. 24, no. 2, pp. 1063–1076, 2018, doi: 10.1109/TVCG.2017.2653117.

[7] I. Jeelani, K. Han, and A. Albert, “Development of virtual reality and stereo-panoramic environments for construction safety training,” *Engineering, Construction and Architectural Management*, vol. 27, no. 8, pp. 1853–1876, Jan. 2020, doi: 10.1108/ECAM-07-2019-0391.

[8] C. B. Rogers, H. El-Mounaryi, T. Wasfy, and J. Satterwhite, “Assessment of STEM e-learning in an immersive virtual reality (VR) environment,” *Computers in Education Journal*, vol. 8, no. 4, Oct. 2017, doi: 10.18260/p.26336.

[9] A. S. Alqahtani, L. F. Daghestani, and L. F. Ibrahim, “Environments and system types of virtual reality technology in STEM: A survey,” *Int J Adv Comput Sci Appl*, vol. 8, no. 6, 2017.

[10] M. Mott, M. Franco, M.G.Holz, C.Ofek, R.Stoakley & M.R.Morris., “Accessible by design: An opportunity for virtual reality,” in *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 2019, pp. 451–454.

[11] S. A. W. Andersen, S. Foghsgaard, P. Cayé-Thomasen, and M. S. Sørensen, “The effect of a distributed virtual reality simulation training program on dissection mastoidectomy performance,” *Otology & Neurotology*, vol. 39, no. 10, pp. 1277–1284, 2018.

[12] N. E. Seymour, “VR to OR: a review of the evidence that virtual reality simulation improves operating room performance,” *World J Surg*, vol. 32, no. 2, pp. 182–188, 2008.

[13] A. Arya, P. Hartwick, S. Graham, and N. Nowlan, “Collaborating through space and time in educational virtual environments: 3 case studies,” *Journal of Interactive Technology and Pedagogy*, vol. 2, 2012.

[14] J. Cecil, P. Ramanathan, and M. Mwavita, “Virtual Learning Environments in engineering and STEM education,” in *2013 IEEE Frontiers in Education Conference (FIE)*, 2013, pp. 502–507.

[15] M. Sauter, D. H. Uttal, D. N. Rapp, M. Downing, and K. Jona, “Getting real: the authenticity of remote labs and simulations for science learning,” *Distance Education*, vol. 34, no. 1, pp. 37–47, 2013.

[16] C.H. Chen and C.H. Chiu, “Employing intergroup competition in multitouch design-based learning to foster student engagement, learning achievement, and creativity,” *Comput Educ*, vol. 103, pp. 99–113, 2016, doi: 10.1016/j.compedu.2016.09.007.

[17] F. P. Brooks, “What’s real about virtual reality?,” *IEEE Comput Graph Appl*, vol. 19, no. 6, pp. 16–27, 1999.

[18] X. Chen, “STEM Attrition: College Students’ Paths into and out of STEM Fields. Statistical Analysis Report. NCES 2014-001.,” *National Center for Education Statistics*, 2013.

[19] R. Raman, H. Vachharajani, and K. Achuthan, “Students motivation for adopting programming contests: Innovation-diffusion perspective,” *Educ Inf Technol (Dordr)*, vol. 23, no. 5, pp. 1919–1932, 2018.

[20] A. Scavarelli, A. Arya, and R. J. Teather, “Virtual reality and augmented reality in social learning spaces: A literature review,” *Virtual Real*, vol. 25, pp. 257–277, 2021.

[21] NSERC/CRSNG, “Guide for Applicants: Considering equity, diversity and inclusion in your application,” https://www.nserc-crsng.gc.ca/NSERC-CRSNG/Policies-Politiques/EDI_guidance-Conseils_EDI_eng.asp 2017.

[22] V. Potkonjak *M.Gardner. V.Callaghan, P.Mattila, C.Gutl*, V.M.Perovic, K.Jovanovic, “Virtual laboratories for education in science, technology, and engineering: A review,” *Comput Educ*, vol. 95, pp. 309–327, 2016.

[23] M-B. Ibáñez and C. Delgado-Kloos, “Augmented reality for STEM learning: A systematic review,” *Comput Educ*, vol. 123, pp. 109–123, 2018.
[24] National Academies of Sciences and Medicine, *Undergraduate research experiences for STEM students: Successes, challenges, and opportunities*. National Academies Press, 2017.
[25] A. Jolly, “Six characteristics of a great STEM lesson,” *Educ Week*, pp. 1–11, 2014.
[26] M. Sanders, “STEM, STEM education, STEMmania. the technology teacher.” Virginia Tech Blacksburg, 2009.
[27] A.P.Carnevale, “STEM 2011,” *Center on Education and the Workforce, Georgetown University.*, 2011.
[28] R.W. Bybee, “What is STEM education?,” *Science (1979)*, vol. 329, no. 5995, p. 996, 2010.
[29] J. Mervis, “Straight Talk about STEM Education,” 2007.
[30] M.Tight, “*Higher Education Research: The Developing Field*. Bloomsbury Academic” , 2018.
[31] C. L. Snelson, “Qualitative and Mixed Methods Social Media Research,” *Int J Qual Methods*, vol. 15, no. 1, 2016.
[32] M. Akçayır and G. Akçayır, “Advantages and challenges associated with augmented reality for education: A systematic review of the literature,” *Educ Res Rev*, vol. 20, pp. 1–11, 2017, doi: https://doi.org/10.1016/j.edurev.2016.11.002.
[33] https://endnote.com/, “EndNote” (accessed Feb. 5, 2019).
[34] https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home, “NVIVO”. (accessed March.10, 2018)
[35] T. R. Kelley and J. G. Knowles, “A conceptual framework for integrated STEM education,” *Int J STEM Educ*, vol. 3, no. 1, pp. 1–11, 2016.
[36] M.Baxter, H. Beauchamp, K.Bouchard, D.Habermas, M.Huff, B.Ladd, J.Pearon, and G.Plague., “Rethinking STEM education: An interdisciplinary STEAM curriculum,” *Procedia Comput Sci*, vol. 20, pp. 541–546, 2013.
[37] D. H. Uttal and C. A. Cohen, “Spatial thinking and STEM education: When, why, and how?,” in *Psychology of learning and motivation*, vol. 57, Elsevier, 2012, pp. 147–181.
[38] J. B. Carroll,"Human cognitive abilities: A survey of factor-analytic studies" . *Cambridge University Press*.1, 1993.
[39] E. Faulconer, Z. Dixon, J. C. Griffith, and H. Frank, “Surveying the Safety Culture of Academic Laboratories.,” *J Coll Sci Teach*, vol. 50, no. 2, 2020.
[40] J. H. Gibson, I. Schröder, and N. L. Wayne, “A research university’s rapid response to a fatal chemistry accident: Safety changes and outcomes,” *J Chem Health Saf*, vol. 21, no. 4, pp. 18–26, 2014.
[41] W.Winn, M.Windschitl, R.Fruland, & Y.Lee, “When Does Immersion in a Virtual Environment Help Students Construct Understanding? ,” *International Conference of the Learning Sciences,* vol. 206, pp. 497–503, Oct. 2002.
[42] M. Harle and M. Towns, “A review of spatial ability literature, its connection to chemistry, and implications for instruction,” *J Chem Educ*, vol. 88, no. 3, pp. 351–360, 2011.
[43] C. K. Baker and T. M. Galanti, “Integrating STEM in elementary classrooms using model-eliciting activities: responsive professional development for mathematics coaches and teachers,” *Int J STEM Educ*, vol. 4, no. 1, pp. 1–15, 2017.
[44] Ş. Purzer, “The relationship between team discourse, self-efficacy, and individual achievement: A sequential mixed-methods study,” *Journal of Engineering Education*, vol. 100, no. 4, pp. 655–679, 2011, doi: 10.1002/j.2168-9830.2011.tb00031.x.
[45] B. Mohamed-Salah and D. Alain, “To what degree does handling concrete molecular models promote the ability to translate and coordinate between 2D and 3D molecular structure representations? A case study with Algerian students,” *Chemistry Education Research and Practice*, vol. 17, no. 4, pp. 862–877, 2016, doi: 10.1039/c5rp00180c.
[46] M. M. Cooper and R. L. Stowe, “Chemistry education research—From personal empiricism to evidence, theory, and informed practice,” *Chem Rev*, vol. 118, no. 12, pp. 6053–6087, 2018.
[47] D. Gabel, “Improving teaching and learning through chemistry education research: A look to the future,” *J Chem Educ*, vol. 76, no. 4, p. 548, 1999.
[48] A. T. Stull, L. Fiorella, M. J. Gainer, and R. E. Mayer, “Using transparent whiteboards to boost learning from online STEM lectures,” *Comput Educ*, vol. 120, pp. 146–159, 2018.
[49] G. Korakakis, E. A. Pavlatou, J. A. Palyvos, and N. Spyrellis, “3D visualization types in multimedia applications for science learning: A case study for 8th grade students in Greece,” *Comput Educ*, vol. 52, no. 2, pp. 390–401, 2009.

[50] N. Hoyek, C. Collet, F. di Rienzo, M. de Almeida, and A. Guillot, "Effectiveness of three-dimensional digital animation in teaching human anatomy in an authentic classroom context," *Anat Sci Educ*, vol. 7, no. 6, pp. 430–437, 2014.

[51] J. B. Elliott, M. Gardner, and M. Alrashidi, "Towards a framework for the design of mixed reality immersive education spaces," *Proceedings of the 2nd European immersive initiative summit, Paris, France*, pp. 63–76, 2012.

[52] J. Martín-Gutiérrez, C. E. Mora, B. Añorbe-Díaz, and A. González-Marrero, "Virtual technologies trends in education," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 2, pp. 469–486, 2017.

[53] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda," *Comput Educ*, vol. 147, p. 103778, 2020.

[54] R. Sj. Baker, S. K. D'Mello, M. M. T. Rodrigo, and A. C. Graesser, "Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments," *Int J Hum Comput Stud*, vol. 68, no. 4, pp. 223–241, 2010.

[55] J. H. Steffen, J. E. Gaskin, T. O. Meservy, J. L. Jenkins, and I. Wolman, "Framework of Affordances for Virtual Reality and Augmented Reality," *Journal of Management Information Systems*, vol. 36, no. 3, pp. 683–729, 2019, doi: 10.1080/07421222.2019.1628877.

[56] D.H.Shin, "The role of affordance in the experience of virtual reality learning: Technological and affective affordances in virtual reality.," *Telematics and Informatics,* vol. 34, no. 8, pp. 1826–1836, 2017.

[57] M.Prensky, "Digital game-based learning," *Computers in Entertainment,* vol. 1, no. 1, 2003.

[58] M. J. Mayo, "Video games: A route to large-scale STEM education?," *Science (1979)*, vol. 323, no. 5910, pp. 79–82, 2009.

[59] J. P. Gee, "What video games have to teach us about learning and literacy," *Computers in Entertainment (CIE)*, vol. 1, no. 1, p. 20, 2003.

[60] R. R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," *Am J Phys*, vol. 66, no. 1, pp. 64–74, 1998.

[61] A. Suh and J. Prophet, "The state of immersive technology research: A literature analysis," *Comput Human Behav*, vol. 86, pp. 77–90, 2018, doi: https://doi.org/10.1016/j.chb.2018.04.019.

[62] R. Al-Azawi, A. Albadi, R. Moghaddas, and J. Westlake, "Exploring the potential of using augmented reality and virtual reality for STEM education," in *International Workshop on Learning Technology for Education in Cloud*, 2019, pp. 36–44.

[63] I. Ghergulescu, T. Lynch, M. Bratu, A. Moldovan, C. H. Muntean, and G. M. Muntean, "STEM education with atomic structure virtual lab for learners with special education needs," *EDULEARN18 Proceedings*, vol. 1, pp. 8747–8752, 2018.

[64] D.Bogusevschi, I.Tal, M.Bratu, B.Gornea, D.Caraman, I. Ghergulescu, & G.M.Muntean, "Water cycle in nature: small-scale STEM education pilot," in *EdMedia+ Innovate Learning*, 2018, pp. 1496–1505.

[65] J. Parong and R. E. Mayer, "Learning science in immersive virtual reality.," *J Educ Psychol*, vol. 110, no. 6, p. 785, 2018.

[66] L.W.Bailey, "New Technology for the Classroom," *Educational Technology and the New World of Persistent Learning*, pp. 1–11, 2019.

[67] L.W.Bailey, "Supporting Learners in the Classroom," *Educational Technology and the New World of Persistent Learning.*, pp. 176–187, 2019.

[68] R. B. Kozma and W. S. Vota, "ICT in developing countries: Policies, implementation, and impact," in *Handbook of research on educational communications and technology*, Springer, 2014, pp. 885–894.

[69] J. Zhao, L. Lin, J. Sun, & Y. Liao, "Using the summarizing strategy to engage learners: Empirical evidence in an immersive virtual reality environment.," *The Asia-Pacific Education Researcher*, vol. 29, pp. 473–482, 2020.

[70] E.Liisalotte, "Immersive virtual reality in STEM: is IVR an effective learning medium and does adding self-explanation after a lesson improve learning outcomes?," *Educational technology research and development*, vol. 1, no. 26, 2022.

[71] A. al Hashemi, "*People with special needs". 2018. doi: 10.13140/RG.2.2.17036.16002.

[72] A. B. Kuyini *et al.*, “Disability Rights Awareness and Inclusive Education Training Manual,” *CEVS8 Ghana, GES Special Education Division & Pathfinders Australia. Australia & Ghana/AB Kuyini, TP Otaah, BM Mohamed, H. Mumin, F. Mashod, RK Alhassan, FK Mahama, A. Zakaria, A. Brennan, T. Wallace, R. Kilbourne*, 2015.
[73] “Disabled World,” www.disabledworld.com (accessed Feb. 09, 2020).
[74] Y. Y. Qian and R. J. Teather, “The eyes don’t have it: an empirical comparison of head-based and eye-based selection in virtual reality,” in *Proceedings of the 5th Symposium on Spatial User Interaction*, 2017, pp. 91–98.
[75] G. C. Burdea, “Virtual rehabilitation–benefits and challenges,” *Methods Inf Med*, vol. 42, no. 05, pp. 519–523, 2003.
[76] S.Chandrashekar, “How VR can transform the way people with disability learn,” *https://www.d2l.com/corporate/blog/gaad-virtual-reality-people-disabilities-learn/.*, Feb. 20, 2020.
[77] Y. Zhao, E. Cutrell, C. Holz, M. R. Morris, E. Ofek, and A. D. Wilson, “Seeingvr: A set of tools to make virtual reality more accessible to people with low vision,” in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–14.
[78] S. Maidenbaum and A. Amedi, “Non-visual virtual interaction: Can Sensory Substitution generically increase the accessibility of Graphical virtual reality to the blind?,” in *2015 3rd IEEE VR International Workshop on Virtual and Augmented Assistive Technology (VAAT)*, 2015, pp. 15–17.
[79] M. A. Sukhai, *C.E.Mohler, T.Doyle, E.Carson, C. Nieder, D.Levy-Pinto, & F.Smith*, “Creating an accessible science laboratory environment for students with disabilities,” *Council of Ontario Universities*, pp. 1–28, 2014.
[80] D. Mendes, F. Relvas, A. Ferreira, and J. Jorge, “The benefits of dof separation in mid-air 3d object manipulation,” in *Proceedings of the 22nd ACM conference on virtual reality software and technology*, 2016, pp. 261–268.
[81] K. Gerling, P. Dickinson, K. Hicks, L. Mason, A. L. Simeone, and K. Spiel, “Virtual reality games for people using wheelchairs,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–11.
[82] J. C. Garcia, M. Marron, J. Ureña, and D. Gualda, “Intelligent wheelchairs: Filling the gap between labs and people,” *Assistive Technology: From Research to Practice: AAATE*, vol. 33, no. 2013, p. 202, 2013.
[83] B. Wong and B. Y. L. Wong, “Learning about learning disabilities”. Academic press, 2004.
[84] P. N. Wilson, N. Foreman, and D. Stanton, “Virtual reality, disability and rehabilitation,” *Disabil Rehabil*, vol. 19, no. 6, pp. 213–220, 1997.
[85] R. B. Lewis, “Assistive technology and learning disabilities: Today’s realities and tomorrow’s promises,” *J Learn Disabil*, vol. 31, no. 1, pp. 16–26, 1998.
[86] K. Menke, J. Beckmann, and P. Weber, “Universal design for learning in augmented and virtual reality trainings,” in *Universal Access Through Inclusive Instructional Design*, Routledge, 2019, pp. 294–304.
[87] B.R.Connell, M. Jones, R.Mace, J.Mueller, A. Mullick, E. Ostroff, E. & G. Vanderheiden, “The Principles of Universal Design - Version 2.0,” 1997. [Online]. Available: design.ncsu.edu/cud
[88] A.F.Siu, “Virtual Reality Without Vision: A Haptic and Auditory White Cane to Navigate Complex Virtual Worlds,” *Association for Computing Machinery,CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020.
[89] F.L. Lopez and V. Sundstedt, “A comparative study of eye tracking and hand controller for aiming tasks in virtual reality,” in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 2019, pp. 1–9.
[90] S. Parsons and P. Mitchell, “The potential of virtual reality in social skills training for people with autistic spectrum disorders,” *Journal of intellectual disability research*, vol. 46, no. 5, pp. 430–443, 2002.
[91] H.Yakobi, “Understanding Barriers to Accessibility.” Ontario: coucil of Ontario University, 2013.
[92] T. L. Jeffs, “Virtual reality and special needs,” *Themes in science and technology education*, vol. 2, no. 1–2, pp. 253–268, 2010.
[93] T. Lannen, D. Brown, and H. Powell, “Control of virtual environments for young people with learning difficulties,” *Disabil Rehabil*, vol. 24, no. 11–12, pp. 578–586, 2002.
[94] AJ Ryan, “Thoughts on Accessibility Issues with VR,” *https://ablegamers.org/thoughts-on-accessibility-and-vr/*, Jan. 2017.
[95] K.U. Phillips, “Virtual Reality Has an Accessibility Problem,” *https://blogs.scientificamerican.com/voices/virtual-reality-has-an-accessibility-problem/*, Jan. 29, 2020.
[96] Samsung, “Relúmĭno,” *https://news.samsung.com/global/relumino*, 2017.

[97] B.M Alain, "To what degree does handling concrete molecular models promote the ability to translate and coordinate between 2D and 3D molecular structure representations? A case study with Algerian students," *The Royal Society of Chemistry*, vol. 17, pp. 862–877, 2016.
[98] J.S. Brown, A. Collins, & P. Duguid, "Situated cognition and the culture of learning. Educational researcher, 18(1), pp.32-42.," *Educational researcher*, pp. 32–42, Jan. 1989.
[99] M. Dunleavy, C. Dede, and R. Mitchell, "Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning," *J Sci Educ Technol*, vol. 18, no. 1, pp. 7–22, 2009.
[100] B. Dalgarno and M. J. W. Lee, "What are the learning affordances of 3-D virtual environments?," *British Journal of Educational Technology*, vol. 41, no. 1, pp. 10–32, 2010.
[101] S. A. Ambrose, M. W. Bridges, M. DiPietro, M. C. Lovett, and M. K. Norman, "How learning works: Seven research-based principles for smart teaching", *John Wiley & Sons*, 2010.
[102] E. von Glasersfeld, "Facts and the self from a constructivist point of view," *Poetics*, vol. 18, no. 4–5, pp. 435–448, 1989.
[103] D. H. Jonassen, "A model for designing constructivist learning environments," in *Proceedings of ICCE97*, 1997, vol. 12, pp. 112–123.
[104] C. J. Chen, "Theoretical bases for using virtual reality in education," *Themes in Science and Technology Education*, vol. 2, no. 1–2, pp. 71–90, 2010.
[105] J. Lowyck, "Bridging learning theories and technology-enhanced environments: A critical appraisal of its history," in *Handbook of research on educational communications and technology*, Springer, 2014, pp. 3–20.
[106] P. A. Ertmer and T. J. Newby, "Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective," *Performance improvement quarterly*, vol. 6, no. 4, pp. 50–72, 1993.
[107] G. van Merriënboer and A. B. H. de Bruin, "Research paradigms and perspectives on learning," in *Handbook of research on educational communications and technology*, Springer, 2014, pp. 21–29.
[108] J. M. Nyachwaya and N. B. Wood, "Evaluation of chemical representations in physical chemistry textbooks," *Chemistry Education Research and Practice*, vol. 15, no. 4, pp. 720–728, 2014.
[109] B. R. Belland, A. E. Walker, N. J. Kim, and M. Lefler, "Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis," *Rev Educ Res*, vol. 87, no. 2, pp. 309–344, 2017.
[110] S. M. Alessi, "Fidelity in the design of instructional simulations.," *Journal of computer-based instruction*, 1988.
[111] A. Kolb and D. Kolb, "Eight important things to know about the experiential learning cycle," *Australian educational leader*, vol. 40, no. 3, pp. 8–14, 2018.
[112] S. Freeman *et al.*, "Active learning increases student performance in science, engineering, and mathematics," *Proceedings of the national academy of sciences*, vol. 111, no. 23, pp. 8410–8415, 2014.
[113] S. Billett, "Guided learning at work," *Journal of Workplace learning*, 2000.
[114] D. Rohrer and H. Pashler, "Recent research on human learning challenges conventional instructional strategies," *Educational Researcher*, vol. 39, no. 5, pp. 406–412, 2010.
[115] L. Long, G. Dabrowski, and A. Grant, "Using 360 Virtual Reality to Make Experiential Learning Accessible to All," 2020, [Online]. Available: https://er.educause.edu/blogs/2020/4/using-360-virtual-reality-to-make-experiential-learning-accessible-to-all
[116] D. Passig, D. Tzuriel, and G. Eshel-Kedmi, "Improving children's cognitive modifiability by dynamic assessment in 3D Immersive Virtual Reality environments," *Comput Educ*, vol. 95, pp. 296–308, 2016.
[117] M. Usoh, E. Catena, S. Arman, and M. Slater, "Using presence questionnaires in reality," *Presence*, vol. 9, no. 5, pp. 497–503, 2000.
[118] M.L. McHugh, "The Chi-square test of independence," *Biochem Med (Zagreb)*, vol. 23, pp. 143–149, 2013.
[119] E. , D. P.A. Pardo-Igúzquiza, "Normality Tests for Spatially Correlated Data," *Mathematical Geology* , vol. 36, pp. 659–681, 2004.
[120] T. Van Hecke, "Power study of Anova versus Kruskal-Wallis test," *Journal of Statistics and Management Systems*, vol. 15, no. 2–3, pp. 241–247, 2012.
[121] "Universal Design for Learning." https://udlguidelines.cast.org/, (accessed Sep. 3, 2020)
[122] J. Campos, B. Keshavarz, "Vection and visually induced motion sickness: how are they related?," *Front Psychol*, vol. 6, no. 472, 2015.
[123] S. D. Long, "Visualizing words and knowledge: Arts of memory from the Agora to the Computer," *https://surface.syr.edu/etd/222/.* , Feb. 11, 2022.

[124] N. Burgess, "Spatial cognition, and the brain," *A. Kingstone & M. B. Miller (Eds.), The year in cognitive neuroscience,Blackwell Publishing*, pp. 77–97, 2008.

[125] T. Dalgleish, "Method-of-Loci as a Mnemonic Device to Facilitate Access to Self-Affirming Personal Memories for Individuals With Depression," *Sage Journals*, vol. 1, no. 2, pp. 156–162, Feb. 2013.

[126] R. N. Carney, "Mnemonic device," *Encyclopedia of Child Behavior and Development , https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-79061-9_1810*, 2011.

[127] S. Mandal, "Brief Introduction of Virtual Reality & its Challenges," *Int J Sci Eng Res*, vol. 4, no. 4, pp. 304–309, Apr. 2013.

[128] E. Krokos, C.Plasiant, A.Varshney "Virtual memory palaces: Immersion AIDS recall - virtual reality," *SpringerLink*, vol. 23, pp. 1–15, Aug. 2018.

[129] J.P. Huttner and R.-B Susanne, "An Immersive Memory Palace: Supporting the Method of Loci with Virtual Reality," *ResearchGate*, Aug. 2017.

[130] H. K.Wu, & P.Shah, P."Exploring visuospatial thinking in chemistry learning". *Science Education*, vol. 88, no. 3, pp.465-492, 2004.

[131] R.Kozma, E.Chin, J.Russell, & N.Marx, N. "The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning." *The Journal of the Learning Sciences* vol.9, no. 2 pp.105-143,2000.

Appendix I. Pilot and Study 2a

A) Pre-experiment survey

1. Participant ID:
2. Age:
3. Handedness:
    a. Left
    b. Right
4. Gender:
    a. Male
    b. Female
    c. Other. Specify:
    d. Prefer not to say
5. How many chemistry courses have you taken as an undergraduate?
6. Have you completed WHMIS safety training before?
    a. Yes
    b. No
7. Which statement is true about Virtual Reality (VR):
    a. I am aware of VR and have experienced it.
        i. Specify your experience type:
    b. I am aware of VR but have never experienced it.
    c. I am not aware of VR and have not experienced it.

==

Knowledge Test

Participants ID: _________

1. Safety in the laboratory is:

- The responsibility of the student only
- The responsibility of the professor only

- A shared responsibility

2.To reduce the chance of accidents:
- Use personal protective equipment (PPE)
- Use the smallest quantity of material necessary
- When possible, replace a less hazardous chemical with a more hazardous one
- All of the above

3. Examples of personal protective equipment do NOT include:
- Goggles and long pants
- Long-sleeve shirts
- Lab coats
- Contact lenses
- All of the above

4. When using gloves as personal protective equipment, which of the following procedures should be followed?
- Inspect gloves for small holes or tears before use
- Remove gloves before handling objects such as doorknobs, telephones, pens, and computer keyboards
- Wear gloves of a material known to be resistant to leakage by the substances in use
- Decontaminate or wash gloves before removing them
- Replace gloves periodically, depending on the frequency of use
- All of the above

5. Which statement is true about Latex gloves:
- Okay to reuse as long as they are clean
- Can be reused only if have not been permeated
- Should never be reused

6. Jewelry is a potential safety issue because:
- It can be damaged by chemical fumes and spills
- It could be stolen
- Chemicals can be trapped under it, in contact with sensitive skin

7. What are the red, green, and black bins for?
- Red:
- Green:
- Black:

8. True or false: Safety regulations require that contact lenses NOT be worn in the lab.
True / False

9. What is the first thing to do if any accident occurs in the lab?

B) Post-experiment knowledge test:

Question 1:
Sophie (Figure 100) is about to enter the lab and run a reaction experiment. What type of Personal Protective Equipment (PPE) is she missing? Point out any other safety issues you see.

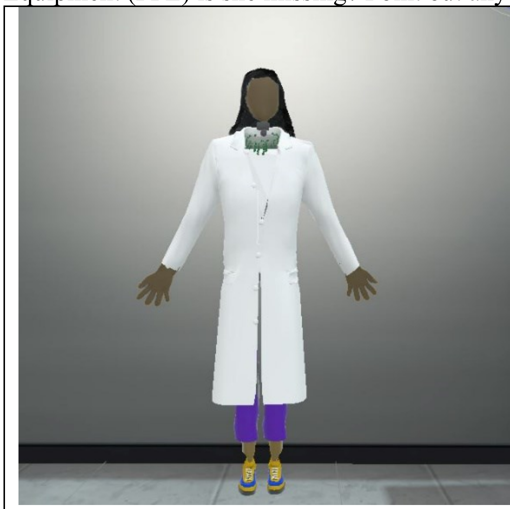
	Correct answers would be: 1-Needs Goggle 2-Gloves 3-Should tie long hair 4-Needs to remove neckless/Jewelry 5-Need long pants to cover the ankle

Figure 71.Illustration used to identify missing PPE

Question2:
You are experimenting with Acid-base (exothermic) reaction using H_2SO_4 and NaOH on the fume hood. What are the right procedure and steps? (Fill in the blank)

Shown to participants	Correct answers
1. Open the fume hood door 2. Put the ice bath inside 3. ______________________ 4. Pure NaOH on it 5. Reaction/ spills on the face happen 6. ____________________	1. Open the fume hood door 2. Put the ice bath inside 3. Put H_2S0_4 in the ice bath 4. Pure NaOH on it 5. Reaction/ spills on the face happen 6. Find the nearest eye washer Use it (push the button), wash your face

C) Usability Questionnaire

Likert Scale from 1 to 7 (7 being the least favorable value)

Rate the systems in terms of:

Ease of use	1- Very easy	7- Very difficult
Memorability	1- I remember very much	7- I remember very little
*Learnability	1-I learned a lot	7-I learned a little
Pleasantness	1-Very pleasing	7-Not pleasing at all
Overall satisfaction	1. Very much	7 - Not at all

**Rapid understanding of the UI/process leads to efficiency in completing the task*

D) Sense of presence survey for IVR (SUS)

Did you have a sense of "being there in the virtual environment"?
2- Where there were times during the experience when the virtual environment was a reality for you?
3- When you think back to the experience, do you think of the virtual environment more as images that you saw or more as somewhere that you visited?
4- During the time of the experience, which was the strongest overall, your sense of being in the virtual environment or of being elsewhere?
5- Do you think a virtual environment as a place in a way similar to other places you've been to today?
6- During the time of your experience, did you often think to yourself that you were actually in the virtual environment?

Any other comments about your experiment in virtual reality?

Appendix II. Study 2b

A) Post-experiment Questionnaire

1. Participant ID: [ ] * You have received this code from within the VR app.
2. Age:
3. Handedness:
    a. Left
    b. Right
4. Gender:
    a. Man
    b. Woman
    c. I prefer not to answer
    d. Something else. Specify if you prefer:
5. Which statement is true about your experience with Virtual Reality (VR):
    a. I am familiar with VR and have experienced it.
        i. Specify your experience type:
    b. I am familiar with VR but have never experienced it myself.
    c. I am not familiar with VR and have not experienced it.

==

B) Knowledge test for all groups (A, B, C)

Question5
a) *Draw (or select) the **Orbital structure** of each compound BeCl2 and annotate/show the correct angle.
b) Draw (or select) the **Lewis structure** of BeCl2 Annotate/show the correct angle.

Question 6
a) Draw (or select) the **Orbital structure** of each compound BF3, annotate/show the correct angle.
b) Draw (or select) the **Lewis structure** of BF3, annotate/show the correct angle.

Question 7:
a) Draw (or select) the **Orbital structure** of each compound NH4, annotate/show the correct angle.
b) Draw (or select) the **Lewis structure** of CH4, annotate/show the correct angle.

Each question has 2 parts, a) for the orbital structure and b) for the Lewis structure
*Considering the limitation of the online-survey tool, students might not be able to "Draw" therefore we might ask the same questions using a multiple-choice question and provide them with 3 answers to choose from. If the experiment is done in person, students will be given a paper version of the knowledge test and will be asked to draw/ write the answers.

Question 8:
Rate the following question from 1 to 5 (1 strongly disagree   5 strongly agree)
This VR visualization helps to understand the SPS concept (angles) better compared to what I read in the textbook.

Question 9:
Which of the following statements is agreeable to you?
a) My test score would improve if I had a chance to complete the VR experiment before taking the knowledge test
b) Completing this VR experiment before the knowledge test would not help me to score better"

C) Usability Questions for VR groups B and C (both interaction types)

Question 1. On a scale of 1-5 (1 very negative and 5 very positive, rate your VR experiment related to each of the following criteria:

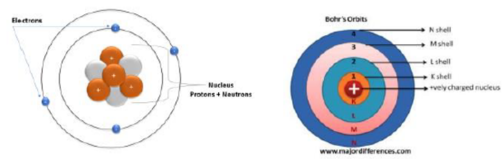
	1	2	3	4	5
Ease of use					
Memorability					
Learnability					
Pleasantness					
Clarity					
Visualization					
Overall satisfaction					

Question 2. Any other comments about your experiment in virtual reality?

D) Text content for 2D group reading

Hybridization and Lewis Structures

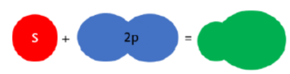
Atoms are made up of electrons, protons, and neutrons. The protons and neutrons are in the nucleus of the atom. The electrons orbit the nucleus in orbitals. An orbital is defined as a region in space in which there is a high probability of finding an electron.



Bonding between atoms:

Bonding happens when two **valence** orbitals overlap and share their electrons to create a sigma bond. In valance-bond theory, electrons of two atoms occupy the same space. This is called overlap of orbitals.

(1S + 1S)



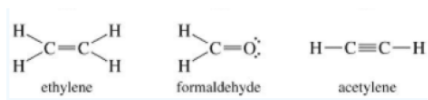
(1S + 2Px)

Hybridization:

Hybridization is the idea that atomic orbitals fuse to form newly hybridized orbitals, which in turn, influences molecular geometry and bonding properties. Hybridization is also an expansion of the valence bond theory. To explore this idea further, we can utilize three types of hydrocarbon compounds to illustrate sp^3, sp^2, and sp hybridization.

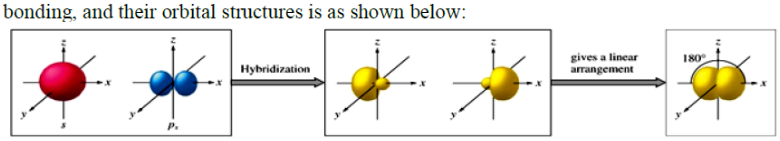
Lewis Structures:

A Lewis Structure is a very simplified representation of the valence shell electrons in a molecule. It is used to show how the electrons are arranged around individual atoms in a molecule. Electrons are shown as "dots" or for bonding electrons as a line between the two atoms. The goal is to obtain the "best" electron configuration, *i.e.* the octet rule and formal charges need to be satisfied.

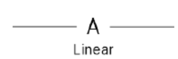


SP hybridization:

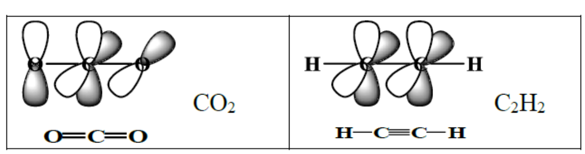
For molecules that have a **linear** (180 degree) geometry, the atoms use sp hybridization for bonding, and their orbital structures is as shown below:



Unhybridized SP hybrid orbitals Linear

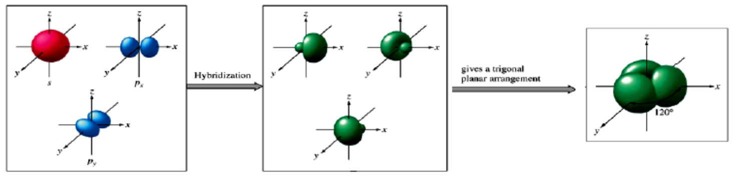


In this hybridization, one s atomic orbital of C combines with one p atomic orbital to create TWO equal energy sp hybrid orbitals lying along the x axis as shown above. Examples of such molecules are CO_2 and C_2H_2. In these carbon-containing sp-hybridized molecules, there are two unhybridized p atomic orbitals of the valence shell of C as shown drawn on the y and z axes, which are used to create the pi bonds in the structures of these molecules.

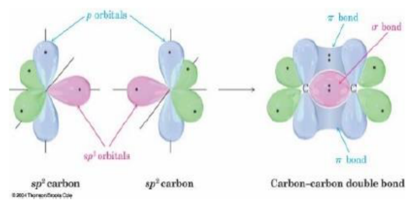


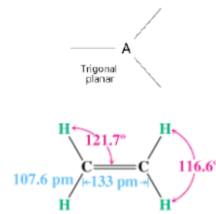
SP^2 hybridization:

For molecules that have a **trigonal planar** (120 degree) geometry, the atoms use sp^2 hybridization for bonding, and the orbital structure is as shown below:



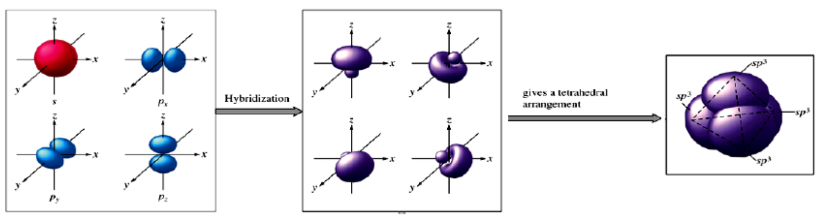
In this hybridization, one s atomic orbital of C combines with two p atomic orbitals of C to create THREE equal energy sp^2 hybrid orbitals lying flat on the plane of a paper as shown above. Example of such a molecule is ethylene (C2H4) shown below, where each C is sp^2 hybridized with one unhybridized p atomic orbital remaining on each C atom (shown in blue) that is used to create the pi bond in the structure:



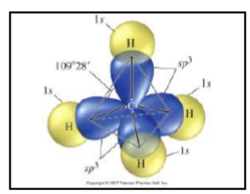


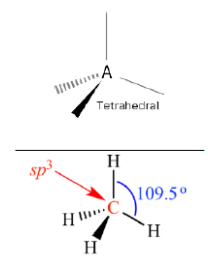
SP^3 hybridization:

For molecules that have a **tetrahedral** (109.5 degree) geometry, the atoms use sp^3 hybridization for bonding, and their orbital structure is as shown below:



In this hybridization, one s atomic orbital of C combines with all three p atomic orbitals of C to create FOUR equal energy sp^3 hybrid orbitals arranged in a tetrahedral shape, as shown above. Example of such a molecule is methane (CH4) shown below, where each C is sp^3 hybridized and there are no remaining unhybridized p orbitals. Thus, it does not contain any pi bonds unlike the sp2 and sp hybridization shown earlier.





Appendix III. Study 3

Post-experiment Questionnaire

1. Participant ID: ☐ * You have received this code from within the VR app.
2. Age:
3. Handedness:
    a. Left
    b. Right
4. Gender:
    a. Man
    b. Woman
    c. I prefer not to answer
    d. Something else. Specify if you prefer:
5. Which statement is true about your experience with Virtual Reality (VR):
    a. I am familiar with VR and have experienced it.
        i. Specify your experience type:
    b. I am familiar with VR but have never experienced it myself.
    c. I am not familiar with VR and have not experienced it.

On a scale of 1-5 (1 strongly disagree …5= Strongly agree), rate your experiment related to each of the following criteria:

6. A- For completing regular tasks (**not requiring raising or lowering your body**) rate your experience:

How **pleasant** was it? (1 Very pleasant 5 Very unpleasant)
1 2 3 4 5
How easy was it **to perform**/complete the task? (1 Very easy 5 Very difficult)
1 2 3 4 5
How easy was it **to learn to perform**/complete the task? (1 Very easy 5 Very difficult)
1 2 3 4 5

7. B- For completing the tasks **(requiring lowering your body/hands)**, rate your experience:

How **pleasant** was it? (1 Very pleasant 5 Very unpleasant)
1 2 3 4 5
How easy was it **to perform**/complete the task? (1 Very easy 5 Very difficult)
1 2 3 4 5
How easy was it **to learn to perform**/complete the task? (1 Very easy 5 Very difficult)

1 2 3 4 5

8. C- For completing the tasks **(requiring raising your body/hands)**, rate your experience:

How **pleasant** was it? (1 Very pleasant 5 Very unpleasant)
1 2 3 4 5
How easy was it **to perform**/complete the task? (1 Very easy 5 Very difficult)
1 2 3 4 5
How easy was it **to learn to perform**/complete the task? (1 Very easy 5 Very difficult)
1 2 3 4 5
==

9. Any other comments about your experiment and accessing objects in the virtual lab?