# Augmented Reality in Chemistry Higher Education

A thesis submitted by

Daniel Elford

in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

UNIVERSITY OF EAST ANGLIA
SCHOOL OF CHEMISTRY

September 2022

# Abstract

Augmented reality (AR) has the capacity to afford virtual experiences that obviate the reliance on using two-dimensional representations of three-dimensional phenomena for teaching chemistry higher education, in addition to positioning students as the protagonists of the learning experience. Thus, the subsequent blending of constructivist pedagogical approaches and AR technology is logical, with this paradigm having enormous methodological potential. Using a combination of quantitative and qualitative instruments, this research project explored the cognitive and affective impacts of engagement with four developed educational interventions, supported using ChemFord, a developed AR application. Firstly, an AR-supported educational escape activity, based on topics of inorganic stereochemistry was constructed. Reported measures of competency were seen as a positive predictor of intrinsic motivation. However, this was not observed to be a positive predictor of academic performance. Next, a Game-Based Learning activity was developed, based on topics of the Valence Shell Electron Pair Repulsion theory. This activity was facilitated both synchronously and asynchronously, exploring the relationships between students' attitudes, perceived cognitive load, spatial ability, and academic performance. Participants demonstrated significant improvements in spatial ability over the study period. In addition, a moderate correlation was found between spatial ability and VSEPR conceptual understanding. The third educational intervention, constructed within a framework of Cognitive Load Theory, illustrates how AR-supported worked examples may enhance learning of electrophilic aromatic substitution. The achievement motivation of learners was also explored, and how this may be impacted by the provision of AR technology and worked examples. Measures of challenge and interest were found to correlate positively with reported germane load, whereas reported extraneous load negatively correlated with measures of challenge and interest for students displaying higher prior relevant chemistry experience. Lastly, a peer instruction session, focusing on topics of coordination chemistry was facilitated. Students' self-efficacy, response switching, and discussions were analysed, in addition to their interactions with the ChemFord application. Students with a lower assessment of their problem solving and science communication abilities were significantly more likely to switch their responses from right-to-wrong than students with a high assessment of those abilities.

# Contents

# List of Figures

**Chapter 4 AR-Supported Gamification for teaching VSEPR**



**Chapter 5 AR-Supported Worked Examples**

**Chapter 6 AR Meets Peer Instruction**

# List of Tables

## Chapter 5 AR-Supported Worked Examples



## Chapter 6 AR Meets Peer Instruction

# Acknowledgements

I wish to thank my supervisors, Professor Simon Lancaster and Doctor Garth Jones for their help and advice during my period of research. In addition, I am grateful to Doctor Anna Wood for insightful discussions concerning the educational intervention described in chapter 6. Lastly, I would like to express my thanks to all students who participated in this body of research.

# Publications

Research undertaken as part of this thesis resulted in the following publications:

Elford, D., Lancaster, S. and Jones, G., 2021. Stereoisomers, Not Stereo Enigmas: A Stereochemistry Escape Activity Incorporating Augmented and Immersive Virtual Reality. *Journal of Chemical Education*, 98(5), pp.1691-1704.

Elford, D., Lancaster, S. and Jones, G., 2022. Exploring the Effect of Augmented Reality on Cognitive Load, Attitude, Spatial Ability, and Stereochemical Perception. *Journal of Science Education and Technology*, 31(3), pp.322-339.

Elford, D., Lancaster, S. and Jones, G., 2022. Fostering Motivation toward Chemistry through Augmented Reality Educational Escape Activities. A Self-Determination Theory Approach. *Journal of Chemical Education*.

Augmented Reality and Worked Examples: Targeting Organic Chemistry Competence. Submitted to Computer and Education: X-Reality.

Peer Instruction meets Augmented Reality. Submitted to Chemistry Education Research and Practice.

# Abbreviations

| | |
|---|---|
| **2D** | Two-dimensional |
| **3D** | Three-dimensional |
| **AIC** | Akaike information criterion |
| **ANOVA** | Analysis of Variance |
| **AO** | Atomic Orbital |
| **AR** | Augmented Reality |
| **ARiEL** | Augmented Reality in Educational Laboratories |
| **ASCI** | Attitude to the study of chemistry instrument |
| **BIC** | Bayesian information criterion |
| **CAEQ** | Chemistry Attitudes and Experiences Questionnaire |
| **CAVE** | Cave Automatic Virtual Environments |
| **CHEMX** | Chemistry Expectations Survey |
| **CI** | Concept Inventory |
| **CLASS** | Colorado Learning Attitudes about Science Survey |
| **CLS** | Cognitive Load Scale |
| **CLT** | Cognitive Load Theory |
| **CT** | ConcepTest |
| **CTML** | Cognitive Theory of Multimedia Learning |
| **CTT** | Classical Test Theory |
| **DFT** | Density Functional Theory |
| **DGBL** | Digital Game-Based Learning |
| **DIF** | Differential Item Functioning |
| **ECL** | Extraneous Cognitive Load |
| **EEA** | Educational Escape Activity |
| **FBX** | Autodesk Filmbox Format |
| **FEHQ** | Frameworks for Higher Education Qualifications |
| **GBL** | Game-Based Learning |
| **GCL** | Germane Cognitive Load |
| **GDPR** | General Data Protection Regulation |
| **GLT** | Generative Learning Theory |
| **GPS** | Global Positioning System |
| **HCI** | Human-Computer Interaction |
| **HF** | Hartree-Fock Method |
| **HMD** | Head-Mounted Display |

| | |
|---|---|
| **HUD** | Head-Up Display |
| **ICC** | Item Characteristic Curve |
| **ICL** | Intrinsic Cognitive Load |
| **ICT** | Information and Communication Technology |
| **iMD-VR** | **I**nteractive Molecular Dynamics in Virtual Reality |
| **IMI** | Intrinsic Motivation Inventory |
| **IRT** | Item Response Theory |
| **ITMC** | Isomerism in Transition Metal Complexes |
| **IUPAC** | International Union of Pure and Applied Chemistry |
| **iVR** | Immersive Virtual Reality |
| **LANL2DZ** | Los Alamos National Laboratory 2 Double-Zeta |
| **LFT** | Ligand Field Theory |
| **MCQ** | Multiple Choice Question |
| **MO** | Molecular Orbital |
| **PDB** | Protein Data Bank |
| **PI** | Peer Instruction |
| **PISE** | Peer Instruction Self-Efficacy Instrument |
| **PSI** | Pilot Study Interviewees |
| **PSVT** | Purdue Visualization of Rotations Test |
| **SCT** | Social Cognitive Theory |
| **SDK** | Software Development Kit |
| **SDT** | Self-Determination Theory |
| **$S_E$Ar** | Electrophilic Aromatic Substitution |
| **SI** | Study Interviewee |
| **SOSESC** | Sources of Self-Efficacy in Science Courses survey |
| **SSQ** | Simulation Sickness Questionnaire |
| **STEM** | Science, Technology, Engineering and Mathematics |
| **STO** | Slater-type orbital |
| **TAM** | Technology Acceptance Model |
| **UEA** | University of East Anglia |
| **UI** | User Interface |
| **UK** | United Kingdom |
| **VR** | Virtual Reality |
| **VRMC** | Virtual Reality Multisensory Classroom |
| **VRML** | Virtual Reality Modelling Language |
| **VRSQ** | VR Sickness Questionnaire |
| **VSEPR** | Valence Shell Electron Pair Repulsion Theory |

# 1

# Introduction

The advancement of information and communication technologies (ICT) has amplified education coverage through digital media, with technologies such as augmented reality (AR) effectively taking root in educational settings. Within chemistry education specifically, resources for visualising chemistry concepts are largely limited to two-dimensional (2D) drawings, static physical models, and pre-programmed animations. Issues with contextualising and visualising abstract chemistry concepts are not only a concern for educators, but also a disconnect from the perspective of students. Visualising the structural changes that molecules undergo throughout a chemical reaction can be a challenging, yet crucial cognitive skill for the inexperienced chemist. Through the adoption of AR, an educator no longer needs to make arbitrary judgements about the most effective representation to carry the learning objective, as accurate, detailed three-dimensional (3D) structures can be instantiated. This technological initiative liberates the 2D constraints of an isometric representation and places control into the fingertips of the learner, promoting active learning in the affective and cognitive domains (An and Holme, 2021; Keller, Rumann, and Habig, 2021). As AR becomes more accessible to educational researchers, the advantages, and pedagogical affordances of integrating immersive technologies into environments that facilitate learning are increasingly reported (Garzón, Pavón, and Baldiris, 2019).

In contrast to immersive Virtual Reality (iVR), AR is a technique that superimposes computer-assisted contextual information on an individual’s view of a real environment (Milgram, Takemura, Utsumi and Kishino, 1994), thus providing a seamless interface for users that blends both the physical and virtual world. This provides the most genuine human-computer interaction (HCI) experience (Cai, Wang, and Chiang, 2014). Reality is not reproduced, but supported, giving students the opportunity to practice their knowledge and skills by combining digital information with the real-world environment (Wojciechowski and Cellary, 2013).

The term was originally coined by Caudell and Mizell (1992) to describe a technology enabling the augmentation of the visual field through utilisation of a heads-up display. This has since been broadened considering that AR can be used to generate multimodal experiences (Akçayır and Akçayır, 2017). Such interactive activities can be based on real-world scenarios. For example, virtual laboratories enable students to play an active role in learning experiences, that may not otherwise be easily replicated due to resources, time, and safety (Chan and Fok, 2009).

## 1.1 ChemFord and Educational Technology

The comparatively low cost of implementing AR technologies into the classroom, on mobile and tablet devices, provides an opportunity for rapid virtual presence, resulting in an accelerated growth in the number of AR educational applications since 2010 (Ozdemir, Sahin, Arcagok, and Demir, 2018). Yet, as a technology in its infancy, AR has several obstacles to surmount before widespread adoption and acceptance. These include issues regarding software usability (Akçayır, Akçayır, Pektaş and Ocak, 2016), resistance from educators (Lee, 2012), overload of cognitive resources (Akçayir and Akçayir 2017; Turan, Meral and Sahin, 2018), and technological limitations (Fraga-Lamas, Fernandez-Carames, Blanco-Novoa and Vilar-Montesinos, 2018; Palmarini, Erkoyuncu, Roy, and Torabmostaedi, 2018). AR systems commonly generate, and sustain, augmented experiences using one, or a combination of, the following approaches:

i. Marker-based.
ii. Markerless.
iii. Location-based.

A marker-based approach extracts natural features from an image, detected using the device's camera, and compares them against a known target resource database (Image Targets | VuforiaLibrary, 2022). These target resources are denoted as image targets and can be any planar image that provides sufficient detail to be detected by an AR system. Once the image target is detected, the AR system will augment the content associated with that image, tracking its position, orientation, and movement. Figure 1.1 shows an example of an image target used with the AR application ChemFord, which was developed as part of this PhD project. ChemFord was developed using Unity editor 2020.1.6f1 and the Vuforia 10.8 software development kit (SDK). ChemFord's design was informed by Cognitive Load Theory

(CLT; Chandler and Sweller, 1991) and minimalism as the underlying design principles.

A markerless system uses a combination of features to determine factors such as the geographic position, or orientation, of a device to sustain an AR experience, eliminating the need for capturing physical targets to trigger virtual interactions. As such, the experience can be easily shared with others, whilst significantly increasing the average range of motion while experiencing AR. However, markerless approaches are mostly dependent on flat, textured surfaces to successfully render virtual objects (Schechter, 2022). Lastly, location-based AR systems utilise Global Positioning System (GPS) and mapping technology to trigger digital content, in place of an image target. When a device approaches a predetermined geographic location, augmented experiences are activated. To support flexibility, ChemFord was developed to utilise both marker-based and markerless approaches. Conditionally, students can scan an image target embedded into a teaching resource to instantiate a particular virtual object. Alternatively, virtual objects can be spawned using ChemFord’s inbuilt inventory, negating the need for an image target.

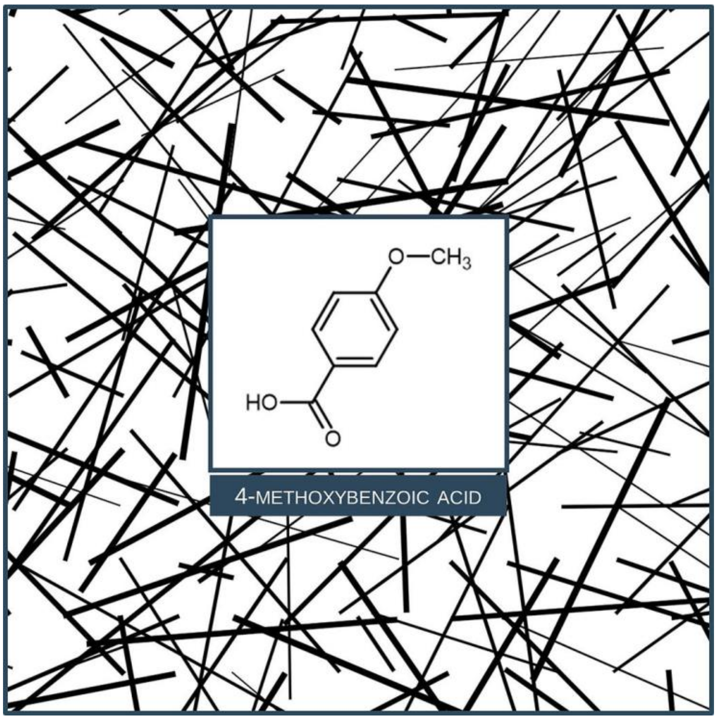



Figure 1.1. A 4-methoxybenzoic acid ChemFord image target.

One of the most famous and enduring debates in the field of educational technology is the debate between Clark (1994) and Kozma (1994). The debate focused on the role of media in the learning process and has been a constant point of contention for decades (Sickel, 2019). Clark's position states that technology, as well as other media, are only vehicles of information. They do not influence student achievement, learning, or motivation, and pedagogical strategies are instead responsible for achieving the purposes of learning. In contrast, Kozma's position states that technology offers unique advantages that enrich learning environments, that could not otherwise be obtained. Whether it is the technology, or the pedagogy, that drives the learning process is still unclear. However, throughout this research study, I posit that AR interventions within education, regardless of the discipline, should consider both the technical affordances of AR, and the pedagogical approach, to generate optimal learning experiences.

Evidence from the literature suggests that constructivism is the most popular learning theory for educational technology (Anderson and Rivera-Vargas, 2020; Duffy and Jonassen, 1992). Like many learning theories, constructivism has been defined and characterised in different ways, and has pedagogical roots associated with the works of Dewey (1938), Vygotsky (1978), and Piaget (1970). What united these psychologists is that all three believed that models such as behaviourism and humanism did not adequately represent the process of learning. Constructivists stress that knowledge is an intersubjective interpretation, and that the construction of knowledge is dependent upon individual and collective understandings (Mascolo and Fischer, 2005). The learner is responsible for constructing their own understanding of the world, using prior knowledge and experiences, to link new information. As such, constructivist learning activities often focus on promoting active inquiry, to encourage learners to develop the skills necessary to effectively solve problems spanning multiple contexts. Four principles of constructivism are outlined as key to understanding its importance:

1. Knowledge is actively constructed by an individual. This will be influenced by learners' prior knowledge and their previous experiences (Schunk, 2020). Practical knowledge is constructed based on the basis of theoretical foundations (Taajamaa, Järvi, Laato, and Holvitie, 2018).
2. Learning is an active process, not a passive one. A learner constructs meaning through active engagement with the environment (Yilmaz, 2008).
3. Learning is a social activity (Palincsar, 1998).
4. Each individual learner will have a unique perspective, based on their existing knowledge and experiences (Fox, 2001). Hence, the same pedagogical approach may result in different learning, as individuals' interpretations of concepts may differ.

As such, educators have adopted several pedagogical approaches derived from these principles of constructivism, to support learners and provide a strong foundation for learning (Saltan and Arslan, 2016; Wen and Looi, 2019). The most common reported in educational interventions that integrate AR technology are:

i. Collaborative learning. Any pedagogical method that advocates or involves groups of learners working together to solve a problem, or to complete a task (Laal and Ghodsi, 2012).
ii. Game-Based Learning. Any type of gameplay with defined learning outcomes (Shaffer, Halverson, Squire, and Gee, 2005).
iii. Inquiry-based learning. The process of learning utilising high-level questioning and exploration to encourage learners to generate real-world connections (Pedaste et al., 2015).
iv. Multimedia learning. A form of media-aided instruction that uses two modalities concurrently (Mayer, 2002).
v. Project-based learning. A student-centered pedagogy where students acquire deeper knowledge through active exploration of real-world challenges and problems (Krajcik and Blumenfeld, 2005).
vi. Situated learning. The acquisition of skills focusing on the relationship between learning and the social situation in which it occurs (Anderson, Reder, and Simon, 1996).

The provision of interactive digital media, such as AR, is a key pedagogical affordance that positions students as the protagonist of the learning experience. Direct manipulation of the properties and relationships of virtual objects that would either be too small, or too large, to examine effectively in a non-virtual environment, can be used to further scaffold the learning of abstract theoretical concepts.

Thus, the subsequent blending of constructivist pedagogical approaches and AR technology is logical, with this paradigm having enormous methodological potential. For applications in chemistry education, embedding virtual experiences has shown to have strong motivational implications. Reported measures have informed that students feel greater levels of motivation when using AR tools, in comparison to other pedagogical tools (Di Serio et al., 2013; Radu, 2012). This may be a direct consequence of sensory engagement, and education researchers are starting to investigate this integration, and its influence, on learning outcomes. Specifically, that the activation of multiple senses enhances knowledge retention (Cheng and Tsai, 2013). Learning gain is the most common reported advantage regarding the use of AR technologies within educational environments (Bacca-Acosta, Fabregat, Baldiris, Graf, and Kinshuk, 2014). Previous research provides quantitative evidence for improvements in academic performance, in addition to qualitative data supporting positive perceptions to its integration. However, it is important to note that these pedagogical benefits are unlikely a consequence of incorporating AR alone, but a combination of different variables that are influenced by AR interventions.

## 1.2 Research Questions

The analysis presented in this thesis was formulated to address a significant gap in the literature; to understand the relationship between the utilisation of AR-supported educational interventions and students' conceptual understanding of chemistry, over a timescale greater than those typically evaluated by cross-sectional studies. The utilisation of cross-sectional studies can make it difficult to make conclusive observations regarding the association between variables. Thus, my research attempts to evaluate how variables such as cognitive and affective factors, in addition to academic performance, change as a function of repeated exposure to a series of developed AR-supported educational interventions. Hence, the nature of this study is longitudinal. The chosen demographic was undergraduate chemistry students enrolled at the University of East Anglia (UEA). The period of research was two academic years commencing in October 2020 and finishing in May 2022. The thesis will focus on three key questions (broken down into sub-questions throughout chapters 3−6) in relation to AR-supported chemistry higher education:

Research Question 1: Do students exposed to the developed AR-supported educational interventions demonstrate improved relevant academic performance in comparison to control conditions?

Research Question 2: What are the cognitive and affective impacts of engagement with the developed AR-supported educational interventions?

Research Question 3: What are the students' perceptions of ChemFord and the developed AR-supported educational interventions as learning experiences?

A mixed-methods approach was employed comprising of quantitative and qualitative methods of data collection and analysis. Throughout chapters 3−6, more detailed accounts of the cognitive and affective factors being investigated, alongside the qualitative insights I sought to capture will be discussed. However, across the four different educational interventions developed, common methods of analysis were employed, which will be discussed throughout the remainder of this section.

Regarding research question 1, the relevant chemistry experience of participants was assessed prior to their engagement with each of the developed interventions. This was done to check whether the experimental groups were significantly different. The data was collected using instruments borrowed from literature, in addition to instruments developed as part of this research study. Where instruments were developed, item and scale characteristics were analysed using procedures of Classical Test Theory (CTT; DeVellis, 2006) and Item Response Theory (IRT; Embretson and Reise, 2000) to gain insight into the overall instrument reliability. After each educational intervention, the same instruments were completed by participants again. This was done to allow learning gain calculations to be completed. Following data collection, data checks (such as normality checks and homogeneity of variances) were conducted to ensure that the correct parametric or non-parametric tests were employed. Normalised change (*c*; Marx and Cummings, 2007) was employed as a measure of the learning gain of students, shown by equation (1.1).

$$c = \begin{cases} \frac{post-pre}{100-pre} & post > pre \\ drop & post = pre = 100 = 0 \\ 0 & post = pre \\ \frac{post-pre}{pre} & post < pre \end{cases} \qquad (1.1)$$

Where pre is a student's pre-test score (%) and post is a student's post-test score (%). Where students score identically at the pre- and post-test stages, *c* values will be 0. Furthermore, students who score 0% or 100% at the pre- and post-test stages are dropped to prevent data from being skewed by uncharacteristically low or high scores respectively. The higher the normalised change, the greater the learning gain. For this study, the ranges defined by Hake (1998) for normalised gain were adopted:

- Low ($c < 0.3$).
- Medium ($0.3 \le c < 0.7$);

- High ($0.7 \le c$).

The properties of equation (1.1) remove the low pre-score bias present with normalised gain calculations. Furthermore, a perfect pre-test score would yield an unbounded normalised gain value, which normalised change avoids by allowing c to take values between −1 and 1. The mean of normalised change $\bar{c}$ is calculated using equation (1.2).

$$\bar{c} = \frac{\Sigma c}{N} \tag{1.2}$$

Qualitative data pertaining to research question 3 was captured through a series of semi-structured interviews with participants over the course of the research period. Interview schedules were constructed around the topics being investigated. Qualitative analysis of participants' interview responses was completed through latent thematic analysis using the approach of Braun and Clarke (2006). Data was recorded, and transcribed verbatim, prior to being subjected to analysis for recurring themes. The initial broad themes were constructed based on the frequency and similarity of responses. Redundancy was eliminated and closely related major themes were merged. I sought to ensure reliability throughout my qualitative analysis using negotiated agreement. Following this, the measure of agreement among coders was calculated using Krippendorff's alpha reliability coefficient. Negotiated agreement extends the interpretation of individual coders into a state of intersubjectivity. Coding differences were discussed and where there was a consistent disagreement, a common approach was agreed. This led to the construction of a set of negotiated codebooks, which were employed throughout analysis. Krippendorff's alpha is a commonly used chance-corrected reliability measure that avoids many of the limitations described for Cohen's kappa, such as its suitability to smaller samples sizes (Krippendorff, 2018). Krippendorff's alpha has ranges between −1.00 and 1.00, with positive values indicating agreement beyond chance. Commonly, data with alpha values ≥ 0.800 are considered reliable, and data with alpha values < 0.800 and ≥ 0.667 are acceptable for tentative conclusions (Krippendorff, 2018).

## 1.3 Epistemological and Ontological Positions

A research paradigm is defined as a "set of common beliefs and agreements" communicated by researchers concerning "how problems should be understood and addressed" (Kuhn, 1962). Researchers tend to adopt one of three research paradigms:

i. Positivist (quantitative, deductive).
ii. Interpretivist (qualitative, inductive).
iii. Pragmatist (mixed-methods).

Each paradigm shapes how an individual seeks to answer research questions and is characterised by its ontological and epistemologist dispositions. As a first approximation, ontology is the study of what there is (Hudson and Ozanne, 1988), whereas epistemology is the relationship between the researcher and the reality, or how we go about capturing this knowledge (Carson, Gilmore, Perry and Gronhaug, 2001).

Pragmatism as a paradigm denotes an inquiry process that is constructed around the combination of qualitative and quantitative methods to answer the research question. A combination of experimentation, surveys, and case studies can be implemented to enhance the quality of the research (Cohen, Manion, and Morrison, 2017). No single method should be relied upon to generate new knowledge. Creswell and Plano Clark (2011) suggest that pragmatism applies to multi-methodological research that draws liberally from both quantitative and qualitative theories, and that researchers are free to dictate the procedures of research that best meet their needs. As such, pragmatism suits this study in relation to the methodological freedom it affords to focus on and answer the research questions. Yet, it is common for a researcher to adopt more than one paradigm. From a pragmatic perspective, I was focused on choosing the research methods that best suited my inquiry. Simultaneously, the interpretivist perspective enabled me to better understand participants' multiple realities.

Researchers of the interpretivist paradigm tend to rely upon participants' views of a studied situation, recognising the impact of their experiences on the research (Creswell et al., 2003). From interviewing participants, to interpreting their views, I am aware that I can never fully understand the meanings that other individuals give to their reality. I can only present my own interpretations of those meanings, and as such, these may be viewed as bias or subjective, even when great caution is exercised to mitigate this wherever possible. The views of participants that engaged with both my educational interventions and ChemFord were explored, to understand how their views may be related to their actions. I wanted to allow new knowledge and themes to emerge, and the results from this data collection, not only provided insight, but also influenced subsequent development of these interventions and tools.

## 1.4 Contribution to the Field

The present research investigates the learning benefits that may be achieved through the utilisation of AR within educational settings in chemistry higher education (Elford, Lancaster, and Jones, 2021, 2022). It makes several noteworthy contributions to the field of chemistry education. Firstly, although ChemFord was not the first chemistry-based AR experience developed, the application itself provides functionality, and thus affordances, such as the dual integration of marker-based and markerless approaches, that I have not identified in other commercially available chemistry-based AR applications. The output of this research also includes four AR-supported educational interventions, where students have displayed statistically significant learning gains following activity engagement.

The Educational Escape Activity (EEA) discussed in **chapter 3** is the first to be used in evaluating students' understanding of the principles of stereochemistry, as well as the first chemistry EEA constructed around the theoretical framework of Self-Determination Theory (SDT). This development also led to the creation of an instrument that may be used to evaluate students conceptual understanding of concepts of stereochemistry. Furthermore, the educational intervention outlined in **chapter 5** are the first AR-supported worked examples utilised within chemistry education. The work conducted utilising this pedagogical approach was also the first to examine achievement motivation in the context of learning electrophilic aromatic substitution. A second instrument was also developed throughout this work, which may be used to assess students' conceptual understanding of electrophilic aromatic substitution. Content validity and reliability analysis were conducted on both developed instruments. This research also provides longitudinal insights into how affective and cognitive factors of participants changed throughout the research period. Lastly, the educational intervention outlined in **chapter 6**, using the pedagogical approach of Peer Instruction (PI) is also the first published using AR-supported discussions for teaching topics of coordination chemistry.

## 1.5 Outline of the Thesis

In this thesis, four AR-supported educational interventions are presented. For each intervention, the pedagogy informing the design is discussed. In addition, I outline how I aimed to afford learning opportunities through the integration of ChemFord. A review of the existing literature is provided in **chapter 2**, beginning with an overview of the challenges faced in chemistry education, in **section 2.1**. An overview of immersive technologies is presented in **section 2.2**, followed by a brief history of AR summarised in **section 2.3**. The pedagogical approaches employed for teaching chemistry with digital media, and more specifically the teaching of higher education chemistry with AR, considering learning theories are outlined in **sections 2.4** and **2.5**, respectively. The remainder of **chapter 2** presents performance metrics, the affordances of AR, and perceptions towards AR technology, in addition to the challenges facing wider implementation of AR technology in educational settings.

In **chapter 3**, the iterative development of an AR-embedded digital EEA, supporting learning of topics of stereochemistry is presented. The design aspects of this intervention were guided by principles of SDT – an intrinsic-extrinsic theory of motivation (Reeve, Deci, and Ryan, 2004). Then, the delivery of an asynchronous online activity, for teaching concepts of valence shell electron pair repulsion theory (VSEPR) is explored in **chapter 4**. This intervention was designed using elements of GBL, and explored the relationships among students' attitudes to study, perceived cognitive load, spatial ability, and academic performance. The introduction of AR-embedded worked examples for teaching concepts of electrophilic aromatic substitution (Hughes-Ingold mechanistic symbol: $S_EAr$) are presented in **chapter 5**. Finally, in **chapter 6**, the integration of AR into a PI session for teaching concepts of organometallic chemistry is described, demonstrating evidence of how conceptual development can occur during AR-supported peer dialogue. An outline of the research timeline is summarised in table 1.1.

Table 1.1. A summary of the educational interventions developed, alongside details of evaluation, throughout the research period.

| Year | Intervention | Chapter | Research variables |
|---|---|---|---|
| 2020 | Pilot EEA (stereochemistry) | 3 | Qualitative interviews, students' experiences |
| | AR gamification (VSEPR) | 4 | Attitude to chemistry study, cognitive load, spatial ability, learning gains, qualitative interviews |
| 2021 | Digital AR-EEA (stereochemistry) | 3 | Motivation, learning gains, qualitative interviews |
| | AR gamification (VSEPR) – repeat | 4 | Attitude to chemistry study, cognitive load, spatial ability, learning gains, qualitative interviews |
| | AR Worked Examples ($S_EAr$) | 5 | Achievement motivation, cognitive load, learning gains, qualitative interviews |
| 2022 | Digital AR-EEA (Stereochemistry) – repeat | 3 | Motivation, learning gains, qualitative interviews |
| | AR-supported PI (Coordination chemistry) | 6 | Self-efficacy, qualitative analysis of students' discussions, ConcepTest response data |

## 1.6 Ethical Considerations

To ensure the feasibility and safety of the selected research topic and recruited participants, ethical guidelines and regulations were considered prior to commencing research to ensure sound research practices. Ethical clearance was obtained under the regulations of UEA's School of Science Research Ethics Committee, a sub-committee of the UEA Research Ethics Committee.

To produce good research, the research topic must have a purpose, while also being relevant and feasible to assess. The purpose of this thesis is to explore university-level chemistry, which is truly relevant to both chemistry education, and modern society. To ensure that the research design was objective, ethical, and coherent, several underlying assumptions were identified:

- The developed educational interventions would be administered using a pre-test/post-test experimental design, to ensure internal validity. Pre-tests can ensure the equivalence of groups, and post-test results can provide an overall effectiveness of the intervention.
- Using a pre-test/post-test design may or may not include control groups. Where control groups are used, participants are randomly assigned between groups.
- The main restriction with this design is that is improves internal validity at the cost of external validity (generalisation).

Educational research involves human subjects, and an individual should, at no point, feel any coercion to participate in a study. As such, participants were informed that their involvement within any aspect of this research was completely voluntary, and that their decision whether to participate would not affect their current, or future, relationships with any individual at UEA. Students must have provided explicit informed consent to participate. The principle of informed consent involves researchers providing sufficient information and assurances regarding the research project to allow individuals to fully understand the implications of participation, to reach a fully informed and freely given decision (DuBois, 2002). This was obtained via a consent form (included as part of the participant information statement). Participants were made aware of their right to withdraw from the study, at any part of the research phase, without declaring a reason. Students were eligible to join the study if they were in the process of completing an undergraduate module in which there was a targeted educational intervention. The students who did not wish to participate in the study were not disadvantaged as the iVR and AR interventions were offered to all students regardless. Students were also able to join the study after it has commenced.

Throughout the research period, participants were assured of data anonymity and confidentiality. Identifying information was irrevocably stripped from data documentation, and study codes utilized in their place. All collected information was used only for the purposes outlined in the participant information statement. Data management was ensured through strict following of data protection principles, outlined under the Data Protection Act 2018, the United Kingdom's (UK) implementation of the General Data Protection Regulation (GDPR). All information was stored securely and only accessible to the researcher. Participant's identities were kept strictly confidential, except as required by law.

The most important ethical consideration is minimizing the risk of harm. Participants should not be harmed in way throughout the research phase. Risks assessments were conducted where necessary, for potential risks to the researcher, participants, and the surroundings. These were clearly defined within the participant information statement. The possible risks of the project were seen as low, and no environmental implications were identified. The research produced from this study is expected to improve the future use of AR technologies for learning in the School of Chemistry at UEA.

# 2

# Literature Review

Details of immersive technologies and their educational application, in addition to the findings, scope and directions of research are presented in this chapter. To start, an introduction to the challenges facing chemistry education is outlined in **section 2.1**. In **section 2.2**, an overview of immersive technologies is presented, followed by a brief history of augmented reality (AR) in **section 2.3**. Next, a general description of the digital media and pedagogical approaches used in chemistry higher education is presented in **section 2.4**, followed by the application of AR in learning environments in **section 2.5**, considering learning theories. Metrics for success in learning environments utilising digital media are outlined in **section 2.6**. This chapter closes with a discussion of the affordances and perceptions of AR in **sections 2.7** and **2.8** respectively, in addition to the barriers facing the wider implementation of AR within educational settings in **section 2.9**.

## 2.1 Challenges in Chemistry Education

Although higher education is fundamentally important to students and can equip them with the skills and knowledge they need to succeed, all of life is an education. Expanding the global community of scientists will be imperative for the future (Rutherford and Ahlgren, 1990); an outlook shared by John Dewey (1910). Over a hundred years ago, Dewey delivered his address ‘Science as a Subject-Matter and as Method’, concluding that the facts of nature are innumerable and inexhaustible. He pleaded for a rethink of what science education should aim to accomplish in shaping the understanding of students. Given that the number of important ‘facts of nature’ has increased by orders of magnitude throughout the last century, this claim continues to be insightful. Science education tends to present the field as ‘an accumulation of ready-made material with which students are to be made familiar’ (Dewey, 1910), rather than an enquiry-based approach. Schwab (1962) denoted this the ‘rhetoric of conclusions’ method to science education. Such an emphasis, Dewey argued, would deprive society of the direction that science could provide. A typical university introductory chemistry module attempts to cover a magnitude of concepts, thus, not enabling the learner to attain either an appreciation, or a deep understanding, of the concepts covered. Hence, Dewey's 'Science Subject Matter' predominates the 'Science as a Method'. Chemistry education must strive to continue experimentation through education research methods, using a diverse range of pedagogical approaches, to determine the relative effectiveness of teaching and learning. AR is one such tool that can be employed for this purpose. Chemistry educators situated in discipline-specific education research are especially well-positioned to drive these innovations. Such faculty members will generally have more resources and freedom to innovate and will be critical for ensuring the execution of well-designed educational experimentation (Wieman, 2017).

The core of theory-driven chemistry education consists of the constant shift between the different representational domains of the macroscopic, the microscopic, and the symbolic. In relation to these concepts, metaphors such as the triangle model developed by Johnstone (1982) – and expanded upon by Mahaffy (2006) – describes how the expert chemist moves between different levels of chemistry (figure 2.1). For students of chemistry, transitioning between these domains is difficult, and educators will hone their language to fit with, and ultimately develop, the students' meaning. The addition of the human element situates chemical concepts, symbolic representations, and chemical processes in the authentic context of human beings (Mahaffy, 2006). This is one of the core challenges in chemistry education (Taber, 2002), and provides a strong rationale for emphasising active learning. This assertion, in addition to the aforementioned bonding misconceptions, contributed to the design of the educational intervention discussed in chapter 6.

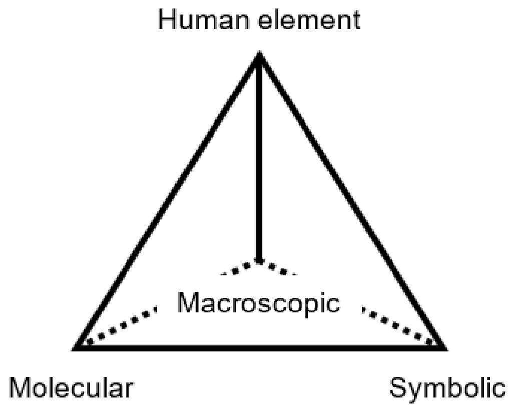



Figure 2.1. The rehybridization of Johnstone's planar triangular metaphor for learning chemistry into Mahaffy’s tetrahedron.

Due to the non-tactile nature of the microscopic domain, educators have sought the integration of AR technology as an initiative to improve the quality and effectiveness of learning environments, whilst enabling and encouraging active learning (Jesionkowska, Wild, and Deval, 2020). Chemistry as a subject is conceptual yet learning chemistry concepts can be difficult (Grove and Lowery Bretz, 2012). Students – whether undergraduate, postgraduate, or at a lower level – often misunderstand; or only partially understand key concepts encountered within their studies. When students demonstrate concepts that are inconsistent with the target knowledge being taught, their ideas may be considered using terms such as misconception, or alternative framework.

Following the constructivist perspective, one of the major sources of students’ preconceptions influencing their learning process are their previous life experiences. Even when educators are aware that an issue is present, these misconceptions can be well established, thus, modifying students’ thinking may not be straightforward. Learners will try to initially apply their prior experiences when constructing new knowledge, regardless of whether fundamentally different concepts must be applied to reach a scientifically reliable understanding (Pfundt, 1982). For example, research has shown that chemical bonding is a difficult concept for students (Luxford and Bretz, 2014; Vrabec and Prokša, 2016). The term ‘covalent bond’ will likely hold meaning for the novice chemist, as they continually encounter the concept within different contexts. This may be restricted to isolated examples, such as homonuclear diatomic molecules (H−H, O−O), or it might be strongly associated with a Lewis structure. However, the expert chemist possesses a far richer meaning for the same term; likely associated with molecular orbitals formed by the linear combination of atomic orbitals, and how this implies information about the physical and chemical properties of a substance. The novice chemist, who has just learnt the notion of covalent bonding, does not yet share the same appreciation as an expert chemist. This is not the case of an educator being correct or the student being incorrect, but of them having different understandings of covalent bonding. Between the two, they use the same word, but not to represent the same thing.

Another considerable barrier inhibiting the learning of essential threshold concepts can be attributed to challenges in visualising spatial concepts such as molecular structure (Barrett and Hegarty, 2016). Issues with contextualising and visualising abstract chemistry concepts is often a concern for educators and is characterised as a disconnect from the perspective of the student. This is widely considered due to a lack of visual observation between students and the molecular domain (Johnstone, 1982). Typically, educators overcome this through the incorporation of 2D drawings. However, the question remains whether a student is capable of successfully forming an accurate objective mental representation of the corresponding three-dimensional (3D) structure. If these four domains, and their interactions are misinterpreted, scientifically unreliable interpretations will likely emerge as a result (Eilks, Möllering, and Valanides, 2007; Johnstone, 1991). Knowledge construction, using immersive technologies, permits learning to be enhanced through the manipulation of these virtual objects. This is afforded through the “transduction of otherwise imperceptible sources of information” (Romli et al., 2015); the observation of areas and events unavailable by other means.

Further explanation through the provision of models and representations are often required to explain the sub-microscopic and representational, due to the non-tactile nature of the domain. The implementation of models is believed to lead to a theory-based understanding at the sub-microscopic level. However, most physical models that are currently employed for chemistry education only provide a limited demonstration of the important aspects of molecules. This is due to their static nature, lacking a connection to the continuous motion, and reactive collisions, which characterises molecules (Grushow and Reeves, 2019). Thus, hindering critical understanding in important chemistry concepts such as thermodynamics. To combat this, new technological paradigms to develop intelligent learning environments, by means of immersive virtual worlds and AR interventions, have been adopted in recent years to move beyond a static representational approach to chemistry higher education (Akçayır, Akçayır, Pektaş, and Ocak, 2016; Griol, Molina and Callejas, 2014). Immersive, experiential learning may surpass traditional strategies (Dede, 2009) as a student's understanding of the spatial arrangement of molecules improves as a result of the manipulation of dynamic, interactive representations of scientific phenomena within an immersive virtual environment.

Related to the affective domain, which includes motivations and attitudes (Bloom, Krathwohl, and Masia, 1973), the introduction of immersive technologies can also support the challenge of improving levels of student motivation when learning about chemistry (Edwards, Bielawski, Prada, and Cheok, 2019; Tüysüz, 2010). This is a substantive factor, with numerous studies commonly reporting that students felt more motivated when using AR technologies, in comparison to other pedagogical tools (Mazzuco, Krassmann, Reategui and Gomes, 2022). This has been shown to inspire productive attention and effort towards learning (Bernholt, Broman, Siebert and Parchmann, 2018). However, it is important to understand how much of this is a direct result of the novelty effect – defined as increased motivation or perceived usability of a technology due to newness (Koch, von Luck, Schwarzer, and Draheim, 2018). This is inherently transitory. As a student becomes familiar with a technology, their engagement, and hence, learning gains, may decrease. It is noticeable that AR technology provides numerous advantages when properly employed, capable of changing paradigms, and overcoming challenges, in the process of teaching chemistry.

## 2.2 Immersive Technology

Two recent focuses of technology-aided instruction have been AR and immersive virtual reality (iVR), due to their capacity to generate immersive and interactive experiences within an educational context. This style of intervention situates immersive technology as a central feature to support learners' goals. An immersive technology "creates the impression that one is participating in a realistic experience via the use of sensory stimuli, narrative, and symbolism" (AECT, 2014). Consequently, an individual is afforded the ability to construct knowledge through direct experiences by conveying an "illusion of nonmediation" (Lombard and Ditton, 1997) between the user and the virtual world. To understand the possible variations and compositions of real and virtual objects, Milgram et al. (1994) proposed the reality-virtuality continuum (figure 2.2), transitioning from an environment composed of solely real objects, to a purely virtual environment. Within this continuum, mixed reality is defined as the blending of real and virtual objects. Of note, is that this version of the continuum is explicitly concerned with visual displays, and without concern for the coherence of the overall experience. Skarbez et al., (2021) argues that the notion of an environment without an observer is incomplete; that technology, content, and user experience must be considered together to adequately describe mixed reality experiences. Sensory conflicts are inherent to conventional virtual environments. Hence, a revision of the continuum was proposed to include an important characteristic of "external" virtual environments, the inability to manipulate interoceptive senses (Skarbez, Smith, and Whitton, 2021). The extension to a "matrix-like" virtual environment, in which sensory conflicts could be avoided, allows stimulation of both interoceptive and exteroceptive senses, and is the only virtual environment that can exist outside of the mixed-reality spectrum. This is illustrated by the discontinuity in the revised continuum.

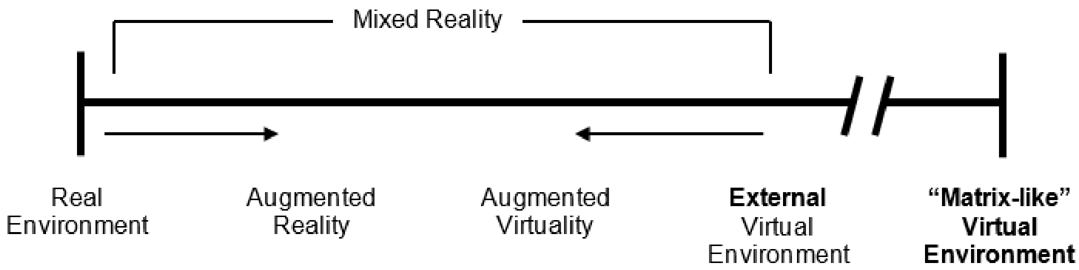



Figure 2.2. Milgram and Kishino's reality-virtuality continuum (adapted from Milgram et al., 1994). Revisions by Skarbez et al. (2021) are shown in bold.

To define the distinction between the terms "interactive" and "immersive", an interactive technology allows for a two-way flow of information through an interface, between the user and the technology. The user communicates a request for data, or action, to the technology, with the technology returning the requested data or action back to the user (IGI Global, 2022, para. 3). Immersion is the provision of sensorial information to an individual that makes them perceive that they are present in a non-physical world. Immersive technologies are generally interactive, however, not all interactive experiences are immersive. It is important to note that immersion should not be confused with presence. Whereas immersion is a technology-related aspect of a virtual environment, presence is a psychological, perceptual, and cognitive consequence of immersion (Mestre and Vercher, 2011). Studies have highlighted the significance of immersion and presence as critical features distinguishing virtual environments from other traditional learning experiences (Fowler, 2015; Kardong-Edgren, Farra, Alinier, and Young, 2019; Makransky and Lilleholt, 2018; Servotte et al., 2020; Shin and Biocca, 2018), and those students who feel more psychologically present can learn in unique manners, afforded through interactions in a multisensory feedback environment (Selzer, Gazcon, and Larrea, 2019).

Multimodal learning refers to teaching concepts using multiple modes. In education, modes are referred to as channels of information that communicate meaning and are experienced by an individual’s senses. As we experience the world, we are provided with data of multiple modalities. These are not mutually exclusive, and will often interact as more information is aggregated, outperforming their unimodal equivalents (Huang et al., 2021). Frequently, iVR and AR developments have focused primarily on visual and auditory senses, to the exclusion of other channels such as haptics, with only a minority of reported works studying the utilisation of haptic technology within immersive environments (Almousa et al., 2019; Imran, Adanir and Khurshid, 2021; Olsson, Nysjo, Seipel, and Carlbom, 2012; Perret and Vander Poorten, 2018). These works comment on the value of touch and tactile sensation as a dimension of immersive environments, observed quantitatively in increased learning scores, and qualitatively in students’ positive responses (Bivall, Ainsworth, and Tibell, 2011). In chemistry education, a tangible virtual environment, which can allow an individual to directly engage with molecular structures and other phenomena; afforded by multimodal channels of learning, could be a powerful tool. In addition, existing literature in the field shows that psychological needs and hedonic experiences (Huang et al, 2018) are an important component of human and virtual world interaction and that virtual environments improve motivation, engagement, and interest. These factors should be included when considering the motivational dynamics of learning experience.

Haptic feedback can be achieved using numerous techniques ranging from desktop pseudo-haptic feedback to human-scale haptic interaction (Richard, Tijou, Richard and Ferrier, 2006). Yet, many researchers do not incorporate haptics into their own developments, thereby failing to leverage the advantages that multisensory experiences may afford for improved learning. Multimodal virtual environments integrate sensory information to create a unified perception (Martin et al., 2022) yielding unprecedented abilities in skill and knowledge transfer, when compared to traditional unimodal learning experiences. This presents an opportunity for potential improvements in learning through increased interaction and greater learner engagement. Tactile feedback is one of the vital ways in which we discover the world around us. It is a key non-verbal means of communication that can promote meaningful learning. Bivall et al. (2011) exploited force feedback, through haptic integration, to allow students to “feel” the forces between interacting molecules.

Post hoc comparisons showed significant intragroup improvements in the haptic condition (13.05, $p < 0.03$) that were not observed in the non-haptic. Yet, care must be taken to ensure that multimodal inputs do not place detrimental cognitive demands on students. When referring to multimodal iVR and AR environments, consideration must be taken to incorporate a sufficient multimodal strategy as the user is not a passive perceiver, but an active learner.

As such, many researchers have studied the use of immersive technologies for their ability to implement context and relationships that are not possible to achieve without the provision of 3D immersive media (Bailenson et al., 2008; Hew and Cheung, 2008; Izatt, Scholberg and Kopper, 2014). Science-based education values the non-threatening, reusable nature of virtual environments. Reusability plays an important role when considering the use of hard to obtain, dangerous, or cost-prohibitive materials (Donally, 2018). They can be quickly adapted to suit an experiment or demonstration, and reduce the time and costs required to modify the corresponding physical experiment. They are not limited to the constraints of the physical environment and can communicate information that would be otherwise invisible. One example is the utilization of virtual surgery training simulations (Sutherland et al., 2006). The introduction of virtual interactive representations has been extensively researched with many studies concluding that greater learning gains may be achieved; especially in subject areas where concepts are abstract or non-representational (Bennie et al., 2019; Edwards et al., 2019; Ferrel et al., 2019; Lupi et al., 2019). Virtual environments can easily provide learners with multiple perspectives and visual cues on the same situation, to make the different aspects of a concept salient.

However, several early reports of virtual experiences have merely taken existing experiences and replicated these traditional approaches in the new environment (Girvan and Savage, 2010). Although this is a natural reaction to the adoption of any new technology, there is a need to move beyond what the technology can replace and consider the unique characteristics that the technology can offer. As such, questions remain unanswered regarding which aspects of virtual environments promote affordances (Shin and Biocca, 2018). This usage of the term affordance was first coined by 1966 by psychologist James J. Gibson as a new perspective on visual stimuli in his book The Senses Considered as Perceptual Systems; and is defined as a characteristic of the environment that, when perceived, affords or provides an opportunity for some action (Gibson and Carmichael, 1966). An affordance enables action on the user who perceives them. In this regard, Gibson conceptualized visual elements of an environment as information, which must be appropriately perceived for the user to recognize the potential of an action. Therefore, the concept of affordances is a pragmatic concept that should guide design decisions that are both functional and easily perceived.

This, in combination with the pedagogical underpinning, is one of the main challenges underlying the development, and utilisation, of immersive technologies for education (Fowler, 2015). Educational technology does not automatically have a positive transformative or empowering on individuals (Southgate and Smith, 2016), and may even distract attention due to the novelty of the environment (Huang, Roscoe, Johnson-Glenberg, and Craig, 2021). Thus, trends of subjective experience or performance, may be partially confounded by novelty effects observed in the predominance of cross-sectional studies. Although most immersive environments are suited for broadly constructivist learning, a lack of structure may result in a 'tyranny of freedom' and would likely place large cognitive demands on an individual. It is important to consider the affordances of immersive technology and explore appropriate pedagogies that could leverage these affordances. Dalgarno and Lee (2010) argue that representational fidelity and learner interaction are two unique characteristics of virtual environments. Representational fidelity refers to the realism of the environment, in terms of sensory feedback and the consistency of object behaviour, whereas learner interaction denotes embodied actions including navigation and the manipulation of those objects, as well as verbal, and non-verbal, communication (Dalgarno and Lee, 2010). Fung et al. (2019) argues that realism not only refers to the visual quality, but also the consistency of object behaviour and available action; and that a lack of visual engagement within an environment deprives students of opportunities to develop critical thinking.

Representational fidelity and learner interaction do not necessarily translate to deeper learning but can afford certain learning opportunities. Reification is one such affordance. Winn and Jackson (1999) used the term 'reification' to represent phenomena that have no natural form, concepts that are abstract and do not correspond directly to material objects. Learners can explore and experiment with concepts such as molecular bonding (Salzman, Dede, Loftin, and Chen, 1999), and radioactivity (Crosier, Cobb and Wilson, 2000) within these virtual microworlds in a style consistent with cognitive constructivist learning theories (Piaget, 1973).

## 2.3 A Brief History of Augmented Reality

The *Pygmalion's Spectacles* is a short piece of literature written by Stanley G. Weinbaum (1935). The story is inspired by Greek mythology where Pygmalion is a sculptor who falls in love with his creation, Galatea, a statue brought to life by Aphrodite, goddess of love. In a similar manner, Weinbaum's writing is embodied by Professor Ludwig, an inventor of a pair of glasses that allows an individual to experience a story in a realistic virtual environment. Although this tale predates computers, it is the very first story to lay the foundations of immersive technologies. In fact, the story anticipates many problems that researchers would inevitably encounter throughout the next century: the high price of hardware, the lack of sociability within an experience, the limited visual quality, and simulation sickness, which is found in the following passage from Weinbaum's (1935) writing:

*"God!" he muttered. He felt shaken, sick, exhausted, with a bitter sense of bereavement, and his head ached fiercely."*

Although early simulation environments date back to the late 1930's with the View-Master, that was patented in 1940 (Bendis, 2003), it was not until the late 1950's that AR began to gain traction. Before its availability to the mass market, AR was establishing itself in the military aircraft sector. In 1958, the first technology supported application of a head-up display (HUD) was manufactured at BAE Systems for the Blackburn "Buccaneer" aircraft (Safi, Chung, and Pradhan, 2019). Notably, AR was utilised inside the cockpit, in the pilot's line of sight, to incorporate digital information as a means of enhancing pilot awareness and safety. Visual clarity was essential, to ensure a pilot's swift comprehension of the real-life exterior environment was not impeded. The projection focal point afforded pilots information – at variable distances – that more accurately matched the depth of the environment, saving them the energy of regularly refocusing, and removing the requirement of a pilot to regularly look 'heads-down' at the instruments. Hence the name 'heads-up'.

Around this time, cinematographer Morton Heilig stated his vision for multi-sensory films in his paper The Cinema of the Future (1955), and proceeded to conceive a detailed design for, what is considered to be, the earliest known example of an immersive, multimodal film experience. By 1962, the first iteration of this design, a mechanical device known as the Sensorama (figure 2.3), had been patented and developed. The Sensorama synchronised many prevalent sensory features such as a stereoscopic 3D display, stereo speakers, and haptic feedback (through vibrations in the user's chair) but was not an interactive experience. Heilig could clearly see the commercial potential for this invention, envisioning training among other applications, but as pioneers often do, he had come too early. There was no adequate technology to support his vision, and thus, no financial backing for further development could be obtained. Heilig also patented the Telesphere Mask, a stereoscopic head-mounted display (HMD) with stereo sound (Heilig, 1960). Both are considered a milestone in the development of immersive technologies.

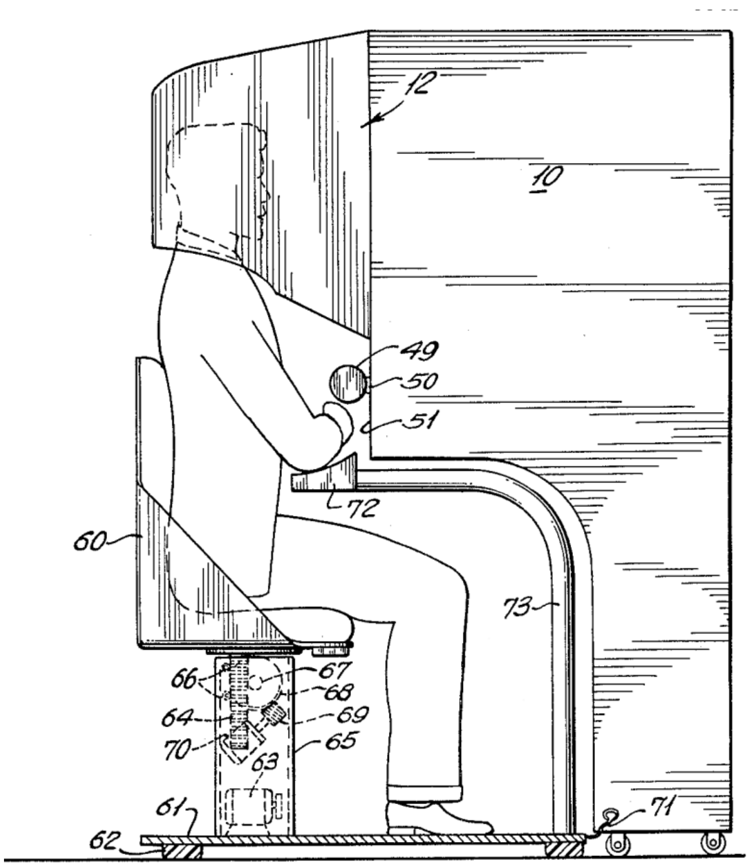



Figure 2.3. Morton Heilig's Sensorama.

Shortly after, Ivan Sutherland (1968), at the University of Utah, published what is believed to be, the first AR HMD. Its name, the *Sword of Damocles*, originates from an ancient Greek parable, and was attributed to the formidable appearance of the device. To perform experiments, a user would have their head securely fastened into the HMD which, due to its weight, had to be attached to a mechanical arm suspended from the ceiling. The HMD was primitive in terms of user interface (UI) and realism, with the virtual environment comprising of simple, computer-generated wireframe rooms (Sutherland, 1968). The system incorporated a stereoscopic display, which would provide different perspectives of the environment depending on the user's head position. The device was partially transparent, so not to disconnect users completely from their surroundings. Consequently, Sutherlands work is often cited as a precursor to modern AR technology.

In the 1970's, Myron W. Krueger (1985) pioneered *Videoplace*, allowing users to interact with virtual objects for the first time. Videoplace consists of two rooms, which can be physically connected, or situated in completely different locations. Each room contains a video camera and special purpose hardware to place users within an interactive environment. On entering a room, a silhouette of the user is projected onto a screen in front of them, in addition to the projection of any users in the other room. All participants observe the same image, and can interact with one another, in addition to virtual objects, using this technology. The movements of users were recorded and transferred to the silhouette representations of users in the environment. Further, either participants' image could be resized and rotated. Through enabling users to perceive the results of their actions on screen, using silhouettes, the users had a sense of presence, even with the absence of direct tactile feedback. The sense of presence was sufficient that users pulled away when their silhouettes intersected with those of other users (Krueger, 1985).

Yet, it wouldn't be until almost 60 years following the publication of Pygmalion's Spectacles that the term augmented reality would be first coined. Tom Caudell (1992), an employee at Boeing Computer Services Research, was asked to create a replacement for Boeing's current wiring instructions used in the production of aircraft. Caudell, and his co-worker, David Mizell, proposed an HMD that superimposed the position of cables through the eyewear, and projected them onto multipurpose, reusable boards. Instead of having to use different boards for each aircraft, the custom wiring instructions could be worn by the workers themselves. Ever since, AR systems have been widely regarded as promising platforms for industrial training and informal learning. In the same year, Louis Rosenburg (1992) developed *Virtual Fixtures* at the United States Air Force.

Virtual Fixtures was first developed as an overlay of virtual sensory information to improve performance in both direct and remotely manipulated tasks (Rosenburg, 1992), and is considered to be the first fully immersive AR system. Due to technological restraints of the decade, 3D graphics could not support the generation of photorealistic and spatially registered AR. Thus, Virtual Fixtures integrated two robots, controlled by a full upper-body exoskeleton worn by the user. A unique optics configuration was employed to generate the immersive experience for the user. This involved the use of a pair of binocular magnifiers, aligned so that the user's perception of the robot arms registered in the same location as their real physical arms. The virtual sensory overlays were presented as either physically realistic structures or abstractions that have properties not possible of real physical structures.

In the context of human-machine collaborative systems, virtual fixtures can be better understood through the provision of an analogy using a physical fixture. A simple task such as drawing a circle is difficult for individuals to perform free-hand with precision and speed. Thus, the use of a physical fixture (a compass) reduces the mental load of the user, improving task performance. The same principle can be applied to a virtual fixture, a task dependent virtual aid that is overlaid upon an environment to provide desired direction on a task, whilst preventing undesired behaviour (Rosenburg, 1992). The result, in terms of Rosenburg's development, was a spatially-registered immersive environment, where the movement of a user's arms, resulted in the simultaneous, and replicated, movement of the robot arms. This was coupled with guides to assist the user whilst performing tasks. Fitts Law performance (Rosenburg, 1992) testing was conducted on users of the system, demonstrating for the first time, that a significant enhancement in human performance of real-world tasks could be achieved by the provision of immersive AR overlays.

Two years later, Azuma and Bishop (1994) developed a motion stabilised display that tackled the main sources of registration error. The system demonstrated accurate static registration across a wide variety of orientations and positions. This was achieved using an optical, transparent HMD (Azuma and Bishop, 1994). In addition, dynamic errors, resulting from the user moving their head position, were reduced by employing inertial sensors to aid head-motion prediction. Azuma also provided future directions in his work titled A Survey of Augmented Reality (1997), which is the most cited article in the field of AR, and one of the most influential MIT Press papers of all time. Azuma is often named one of AR's most recognised experts and is considered to provide a commonly accepted definition for AR as a system (one that has the following three characteristics):

i. Combines the real and the virtual.
ii. Interactive in real time.
iii. Registered in 3D.

With advancements in computer technology came the wider application of AR developments within teaching and learning. Inkpen (1997) studied the effects of learning using computer-based technologies, reporting increased motivation, and learning, of participants engaging with interactive devices collaboratively in a multi-user environment. The turn of the millennia would see the release of ARToolKit (Kato and Billinghurst, 1999), created by Hirokazu Kato, the first open-source software library for AR, allowing widespread development of educational applications. As AR began to find immediate applications in science and math, evaluation of this technology also became imperative. Kaufmann (2003) discusses the utilisation of an evaluation tool known as Construct3D to study the importance of AR in individual and collaborative settings in mathematics. The results of the study confirmed that the benefits of AR are equally present in face-to-face and distance settings (Coimbra et al., 2015). In addition, through utilisation of NeuroEdu, (Xiao et al., 2016) brainwave analysis results indicate that learners' attention and emotional indicators increased while using AR technology, with qualitative interviews revealing higher student satisfaction.

Research into the potential benefits of AR in learning and teaching has steadily grown, with a significant increase in the number of published articles after 2013 (Akçayır and Akçayır, 2017; Dey et al., 2018). Simultaneously, the development race for new AR hardware devices and interfaces commenced. In 2014, Google announced the shipment of Google Glass for customers, subsequently followed by Magic Leap, Microsoft HoloLens, and Meta 2 releasing their developer kits. A Google Scholar-based search for "augmented reality learning" provides almost 1,110,000 results. Although this search does not fulfil the requirements outlined in systematic reviews published within the research field (Buchner, Buntins and Kerres, 2022; Theodoropoulos and Lepouras, 2021), it demonstrates the vast amount of ongoing research and available information.

## 2.4 Teaching Chemistry with Digital Media

Chemistry has the distinction of being a science with iconography, with periodic tables commonplace in most chemistry classrooms and laboratories. Yet, the 3D, dynamic, nature of chemistry are two qualities that are difficult to represent within a textbook or communicate verbally in a lecture. As such, due to the perception that new technologies can support competency to cognitively process multiple external representations (Habig, 2020; Martín-Gutiérrez et al., 2017), there has been a longstanding societal push to adopt multimedia for teaching chemistry. Within this section, the use of such multimedia in chemistry education, excluding AR – which is explored in section 2.5 – is discussed, alongside the pedagogies supporting its implementation.

One example of the remarkable emerging technologies adopted in teaching, and pedagogical research, is the utilisation of Cave Automatic Virtual Environments (CAVE; de Back et al., 2020; Limniou et al., 2008; Lupi et al., 2019). The CAVE was developed to overcome many of the limitations of HMDs, using large, fixed screens positioned in proximity to the user (figure 2.4), thus, minimising the encumbrances worn by participants. In addition, CAVE allows multiple users to share the same experience. The virtual experience is typically viewed in stereoscopic 3D using stereo shutter glasses, which are comprised of polarised lenses (Imai et al., 2000). Learners can be immersed across the full human field of view, unrestricted by an HMD. One advantage of this is that increased field of view has been reported to aid memory (Lin, Duh, Parker, Abi-Rached, and Furness, 2002; Ragan et al., 2010). The implementation of CAVE affords several features that make them uniquely positioned for immersive collaborative learning, by enabling users to simultaneously view their physical body, that of others, and the virtual environment (de Back et al., 2020), as well as mediating several learning benefits including spatial, experiential, and contextual learning (Dalgarno and Lee, 2010). This is dissimilar to iVR experiences, which isolate the user from their physical surroundings, whilst only tracking their head and hand positions. Unlike a CAVE experience, any other body information is consequently lost in an iVR setting or must be inferred by the system.

Figure 2.4. The CAVE system. The standard configuration features three large rear-projection walls and a front projected floor.

Dalgarno and Lee (2010) presented a comprehensive model (introduced in section 2.2) encapsulating most of the aforementioned benefits, referring to theories including Cognitive Load Theory (CLT) and social constructivism. The model details how pedagogical benefits may indirectly arise from the unique characteristics of virtual learning environments by affording learning tasks. While the model of Dalgarno and Lee (2010) is comprehensive, it does not contain examples of the technical elements which representational fidelity and learner interaction consist of. de Back et al. (2020) comment that adding these elements facilitates a comparison of the differential ways in which 2D monitors, VR headsets and CAVEs may afford learning tasks. In their elaborated model, it is argued that configuration of the deployed elements (such as field of view, spatial audio, and tactile force feedback), and characteristics of the task, impact how learning benefits may beneficially occur. This suggests that the benefits of virtual environments are not automatic and require informed design choices to be obtained successfully. Individual differentiating elements do not map directly to learning benefits (de Back et al., 2020). From the model, it can be understood how, through afforded tasks, collaborative learning in CAVEs may benefit from increased embodiment, and the higher level of expressiveness it enables. CAVE allows for natural group interaction, thus creating a strong sense of co-presence (de Back et al., 2020).

Within chemistry education, Limniou, Roberts, and Papadopoulos (2008) report work, in which a CAVE condition for visualising the structural changes of methyl orange, when reacting with an acid, and the formation of acid rain, was compared to the same topic in an equivalent 2D desktop monitor condition. The animations were designed following aspects of CLT. In both conditions, teacher instruction was used, and students were given the opportunity to ask questions. Students' conceptual understanding was assessed using multiple-choice questions (MCQ), with higher learning outcomes reported in the CAVE condition (statistically significant differences found through ANOVA analysis when compared to the 2D desktop condition). Students participating in the CAVE condition were also noted to be more enthusiastic. However, collected student views were not further investigated through qualitative analysis. Additional work by Lupi et al. (2019) describes a CAVE-based virtual laboratory – coupling Caffeine (a molecular viewer) with the natural orbital for chemical valence/charge-displacement scheme - for immersive analysis of chemical bonding, in an interactive and cooperative manner. However, the study focuses on the computational details of development, and does not provide quantitative or qualitative data regarding advantages or affordances of use within an educational setting. In sum, in the few chemistry education CAVE studies published, benefits of CAVE conditions over conventional learning methods were obtained for those studies reporting such differences (Limniou, Roberts, and Papadopoulos, 2008).

A further prominent program for working with molecules in VR is NarupaXR (O'Connor et al., 2018, 2019; Pereshivkina et al., 2021), a framework developed to interface the HTC Vive with rigorous real-time molecular simulation algorithms (iMD-VR). The approach affords multiple optically tracked individuals the ability to manipulate real-time molecular dynamics. Users can "grab" and "pass" molecules back and forth as a result of the interaction site in 3D physical space being exactly the interaction site in 3D simulation space (O'Connor et al., 2019). The simulation is perfectly co-located. The interaction, in which multiple users in the same room can easily pass a simulated molecule between themselves, as if it were a tangible object, represents a class of simulated virtual experience which is not possible within the large-scale immersive stereoscopic CAVE environments (O'Connor, 2018). As each VR client has access to global position data of all other users, any user can see through his/her headset a co-located visual representation of all other users at the same time. As of 2019, NarupaXR supports co-location of six users in the same room, within the same simulation. O'Connor (2018) explored the effectiveness of iMD-VR through recruiting participants to complete three tasks:

i. Threading $CH_4$ through a nanotube.
ii. Changing the screw sense of a helicene molecule.
iii. Tying a knot in a polypeptide (17-ALA).

Qualitative data analysed through thematic analysis indicated an overwhelming preference for the VR system. This was refined into three high-level themes. The first was the impact of depth perception, which was considered important for comprehending the 3D shape of molecules. Secondly, participants positively perceived the control over perspectives, afforded by the capability to freely make molecular manipulations in any dimension, allowing a significantly higher degree of control. Lastly two-handed gesture and sense of agency directly supported participants' gestures for VR molecular interactions. For non-VR platforms, participants are forced to translate physical gestures into a secondary set of gestures adapted to the limitations of the platform. These are far less intuitive when accomplishing complex 3D tasks. Quantitative benefits were also demonstrated in the form of researchers being able to complete molecular modelling tasks more quickly than they can using conventional interfaces such as a mouse or touchscreen (O'Connor, 2019).

Edwards et al. (2019) reports the evaluation of a developed organic chemistry VR Multisensory Classroom (VRMC) as an immersive learning environment, where learners employ hand movements to build hydrocarbon molecules, and experience haptic feedback through gloves with built-in sensors. Although other works employing haptics in VR for chemistry education are documented (Sato et al., 2008), there are few that use direct hand manipulations (Wu et al., 2020) as an alternative to hardware. One further example is Molecular Rift (Norrby et al., 2015), which simulates an advanced organic chemistry classroom for molecular visualisation in drug design. The design of the VRMC is based on the significance and focus of experiential learning theory, as described by Kolb (2014), and consists of an Android phone in a VR headset. Quantitative data collection assessed the VRMC usability as an instructional tool based on its support for multisensory learning, haptics and motivation, which were positively perceived. This theme was also found throughout qualitative feedback, in that participants found the system to be educative and enjoyable. However, some limitations noted include the need for built-in-audio for feedback and instruction, in addition to improved sensitivity and precision of the haptic system.

Based on interest theory (Harackiewicz, Smith and Priniski, 2016), it can be predicted that students who learn in iVR would report greater interest and motivation, and therefore, may score higher on post-tests when employed as part of an experimental design. Parong and Mayer (2018) argue that, based on the Cognitive Theory of Multimedia Learning (CTML), students who learn science with a well-designed slideshow should score higher on a post-test, but will not report higher levels of interest and motivation. To minimise the inherent differences between the two media, slideshows were constructed from the VR lesson to equate the lessons as much as possible, including the narration. Interestingly, the slideshow group scored significantly better than the VR group on the post-test ($p$ = .003, d = 0.92; Parong and Mayer, 2018). Parong and Mayer (2018) conclude that this may have been due to three possibilities: the coherence principle, the segmentation principle, or higher learner control. In addition, segmenting the content of the VR lesson, with written summaries, resulted in greater post-test scores than the original VR lesson. This provides evidence for generative learning theory, that engaging learners both physically and cognitively promotes meaningful learning, and that summarising may result in enhanced learning gains from a lesson. This encouraged learners to select, organise and integrate the information from the VR lesson into their existing knowledge structures (Parong and Mayer, 2018). This lends to the argument that, because no single media attribute can contribute a unique cognitive effect on a learning task, other variables may be more instrumental in learning gains.

## 2.5 Teaching Chemistry with Augmented Reality

The proliferation of AR technologies into learning environments is creating opportunities in concern with offering information to the sensory channels of higher education chemistry students. Numerous benefits have been observed when AR is utilised as a tool to support pedagogy (Pribeanu et al., 2017). The implementation of AR provides educators with new approaches for presenting learning materials (Cai et al., 2021). For example, image targets can be easily combined with print media to support the 3D visualisation of chemistry phenomena, in proximity to relevant textual information. This is beneficial in terms of learning with multimedia (contiguity principle, Mayer and Fiorella, 2014). As such, studies in chemistry education have presented the positive impacts and affordances of AR in terms of increased conceptual understanding – contextual visualisation (Virata and Castro, 2019), learning of spatial structures (Fatemah et al., 2020), information retention (Irwansyah et al., 2020) – soft skills, and motivation (Acosta et al., 2019). In addition, research has shown that AR can foster cognitive and affective learning outcomes of students (Abd Majid and Abd Majid, 2018; Cheng and Tsai, 2013; Olim and Nisi, 2020). In the cognitive domain, the integration of technology that renders the microscopic is of particular interest, and the potential of AR is evident when exploring its utilisation to support learning in the field of chemistry. As such, it is relevant to identify in which topics of chemistry higher education AR has been applied, in addition to the technology employed, the observed advantages, and the challenges encountered.

In their review of the literature, Sırakaya and Alsancak Sırakaya (2020) analysed the target groups in which AR studies were conducted in STEM fields. The results indicated that only 17% were carried out at university-level, with a similar representation being reported in a review by Ibáñez and Delgado-Kloos (2018, 25%). For the studies analysed, reported sample sizes were commonly between 31 and 100 participants (Ibáñez and Delgado-Kloos, 2018; Sırakaya and Alsancak Sırakaya, 2020). In line with a further analysis by Bacca et al. (2014), the most common collection tools used in AR studies were achievement tests (~30%), surveys (~25%), and interviews (~20%). Chen et al. (2017) highlight that the most common research methodology applied to AR studies are mixed methods (40%), of which almost a third utilise a pre-test/post-test experimental design, supported by qualitative data collection.

As a tool for chemistry education, AR appears beneficial regarding the reduction of cognitive load (Buchner, Buntins and Kerres, 2022), and the improvement of spatial performance measures (Rahmawati, Dianhar, and Arifin, 2021). The understanding of abstract concepts in chemistry requires comprehension of phenomena which are inaccessible to sensory experiences. Consequently, they demand high cognitive and spatial capacity, as well as abstraction (Frevert and Di Fuccia, 2019). Considering Bloom’s Taxonomy (1956), the cognitive domain involves knowledge and the development of intellectual skills. Buchner et al. (2022) found that 63% of studies comparing AR versus non-AR conditions displayed lower, or equal, cognitive load, coupled with an increase in learner performance. When comparing AR versus other media, this result was slightly lower at 56% (Buchner, Buntins and Kerres, 2022). These results appear to translate to chemistry education with studies implementing AR reporting lower cognitive load (Rahmawati, Dianhar, and Arifin, 2021), improved performance (An and Holme, 2021), and enhancements in spatial ability (Kodiyah et al., 2020). Just one study compared one type of AR with another. In a study by Chen et al. (2009), students learned organic chemistry using AR glasses or a webcam-AR interface. No differences were observed between the two groups in terms of learning outcomes, or cognitive load. However, these studies are not free from criticism. Studies in educational technology should ensure that all experimental groups are exposed to the same instructional method and content, when comparing media (Mayer, 2019).

Revisiting Bloom’s Taxonomy, the affective domain includes how people deal with things emotionally, such as feelings, values, appreciation, motivations, and attitudes. In their systematic review, Ibáñez and Delgado-Kloos (2018) found that 53% of AR studies investigated the effects of AR learning within the affective domain. For research in chemistry higher education, reported measures of motivation are cited in table 2.1. This aligns to the work of Garzón et al (2019), who found that motivation was the second most reported advantage. For numerous studies, higher measures of motivation to learn skills and knowledge in depth, were reported for immersive and situated AR environments in comparison to other pedagogical tools (Bujak et al., 2013; Chen and Tsai, 2013; Wojciechowski and Cellary, 2013).

Interest is a motivational factor and considered to be a precondition of intrinsic motivation (Fonseca et al., 2014; Renninger and Hidi, 2011; Swarat et al., 2012). Immersion enables students to leverage their interest by increasing their attention and engagement (Bujak et al., 2013; Di Serio et al., 2013; Fonseca et al., 2014). As noted by Renninger and Hidi (2011), interest makes motivation and engagement meaningful. Therefore, it is important to understand how different pedagogical approaches incorporating AR influence chemistry learning outcomes. Many studies do not consider different teaching formats under the same environment. These traditional approaches (Chang and Hwang, 2018; Zhang et al., 2015) do not consider situational interest, an early and temporal phase of interest development. This forms the basis of repeated engagement for knowledge building (Linnenbrink-Garcia et al., 2010; Renninger and Hidi, 2011).

Systematic reviews (Akçayır and Akçayır, 2017; Mazzuco et al., 2022) highlight that smartphone and tablet devices are the most common present in chemistry higher education AR studies (~60%). Qiao et al. (2019) comment on how portable technology now has the hardware requirements to make AR practical for educational integration. In comparison, desktop devices were identified in just 16% of chemistry education AR studies, such as works by Cai et al. (2014) and Maier and Klinker (2013). This observation supports the findings of Alseadoon et al. (2021) who discusses the continual growth in the migration towards mobile platforms that support portability, context-sensitivity, and enhanced usability. It should be noted that mobile and tablet devices offer further unique educational affordances (Klopfer, Squire and Jenkins, 2002) such as connectivity – the ability to connect devices to a common network that creates a truly shared environment, and social interactivity - the ability to exchange data and collaborate in the same physical location. Chen et al. (2017) highlight that the approach of image-based AR is preferred in research studies to generate augmented experiences, compared to markerless and location-based approaches, with the study by Zhu et al. (2018) the only one to utilise the Microsoft HoloLens for chemistry instruction.

AR technology utilises software development kits (SDK) to generate chemistry-based visualisations, with studies examining the potential of programs such as ARChemEx, ARKimia Kit, AC, Vuforia, and ARchemy. The technological evolution of AR, using devices such as HoloLens, affords interactions with virtual objects through digital monitoring of the hands (Grandi et al., 2018). As such, researchers have compared Microsoft HoloLens AR glasses with visualisation experiences on 2D desktop devices. The results of these studies indicated that the 2D teaching approaches were faster and more precise. Consequently, this weakens the technological acceptance of AR technology. A big change driven by AR will be the human interaction with reality, enhanced with digital components. How interactions and recognition of user gestures evolves, alongside the emerging technology, will require rigorous academic and scientific analysis, particularly in the field of education (Fombona-Pascual, Fombona, Vicente, 2022).

Table 2.1. A summary of AR studies in chemistry higher education.

| Author (Year) | Scope of Application | Study Focus | Pedagogical Approach | Outcomes and Major Findings |
|---|---|---|---|---|
| Abdinejad et al. (2021) | Organic chemistry | Students' learning motivation and the usefulness of AR as a learning resource. | Multimedia-Based Learning | AR positively impacted students' learning motivation across multiple constructs: *Attention, Relevance, Confidence, and Satisfaction*. |
| An and Holme (2021) | General chemistry experiments | Learning practical chemistry based on a selected experiment (pH measurement). Students' attitudes were also measured. | Web-Based Learning | Remarkable differences were observed between the pre- and post-test surveys. A 4% increase was reported on the intellectual accessibility sub-scale alongside a 6% decrease on anxiety scores. |
| Aw et al. (2020) | Visualisation of molecules | Investigating learning experiences integrating the visualisation of molecular structures | Multimedia-Based Learning | 45% of students felt more confident learning nucleophilic addition mechanisms after using NuPOV. |
| Erikson et al. (2020) | Molecular visualisation | To provide a workflow for visualising 3D structures in an AR environment. | Computer-Based Learning | Description of a workflow for visualising the 3D structures of molecular systems. |
| Estudante and Dietrich (2020) | Solvay process | Implementation of AR into a digital chemistry escape room activity. | Inquiry/Game-Based Learning | 96% of respondents agreed that the AR activity was a good tool for increasing motivation. |
| Habig (2020) | Stereochemistry | Impact of gender on spatial ability and attitudes to study. | Not explicitly stated | Males scored higher than females on AR-based items. No significant differences between genders on attitude to study. |

| Author (Year) | Scope of Application | Study Focus | Pedagogical Approach | Outcomes and Major Findings |
|---|---|---|---|---|
| Irwansyah et al. (2018) | Molecular geometry | Development of an Android-based AR tool for learning molecular geometry. | Not explicitly stated | Qualitative feedback on the AR application scored highly on motivation, learning objective relevance, and learning support capacity. |
| Kodiyah et al. (2020) | Conformation of alkanes and cycloalkanes | Effect of AR media on improvement in spatial ability. | Multimedia-Based Learning | A moderate N-gain of 0.58 was reported for the association of AR media and the enhancement of spatial ability. |
| Lu et al. (2021) | Real-life chemistry | Students' perceptions on the use of novel AR software to support flipped and gamified learning | Game-Based Learning | Students' perceptions were assessed using a 6-point Likert scale. The highest score of 4.72 was obtained for cognitive accessibility, while cognitive validity yielded the lowest score at 4.01. |
| Naese et al. (2019) | Analytical instrumentation | To enable students to perceive how analytical instruments in a lab operate. | Web-Based Learning | Participating students liked using AR to learn more about the instruments used within the lab. |
| Nazar et al. (2020) | Molecular geometry | To measure the usability of AR for learning molecular geometry. | Not explicitly stated | The application was considered useful for learning molecular geometry. |
| Núñez et al. (2008) | Inorganic chemistry and Crystal Structures | To improve students' understanding of material structures. | Collaborative Learning | 40% of students agreed that the main advantage of AR was 3D interaction. 20% stated that the approach was valuable for the development of spatial skills. |

| Author (Year) | Scope of Application | Study Focus | Pedagogical Approach | Outcomes and Major Findings |
|---|---|---|---|---|
| Plunkett (2019) | Molecular visualisation | To provide a workflow for visualising 3D structures using AR. | Multimedia-Based Learning | Description of a workflow for visualising the 3D structures of molecular systems. |
| Rubilar (2019) | Contextualised chemistry topics | To offer possible pointers regarding AR in chemistry throughout the teaching and learning process. | Multimedia-Based Learning | AR can be used to improve visualisation processes. |
| Tee et al. (2018) | Colorimetric titration | The demonstration of an AR colorimetric titration tool that operates using students' devices. | Multimedia-Based Learning | The tool was effective in helping students attain good bench skills with little use of liquid chemical, thus reducing handling risks and environmental impacts. |
| Wulandari et al. (2019) | Molecular geometry | The application of AR media for teaching concepts of molecular geometry. | Cooperative Learning | Results showed good performances of participants as a result of AR media and discussion. |
| Yang et al. (2018) | Real-life chemistry | Teachers' perceptions of an AR application for chemistry education called "Elements 4D". | Inquiry-Based Learning | Content analysis revealed that participants generally had a positive attitude towards AR as a chemistry learning experience. |
| Zhu et al. (2018) | Biochemistry and lab safety | To develop an AR program to enhance the learning experience of laboratory safety using Microsoft HoloLens. | Inquiry-Based Learning | Students who used the AR condition recalled the location of more items in the laboratory compared to those in the control condition. |

Within the research field of AR in chemistry higher education, there appears to be five major topics where AR research has the highest reported occurrences (Mazzuco et al., 2022). The most reported topic is that pertaining to molecular structures (~19%), such as work by Abdinejad et al. (2021). This is followed by the topics of chemical reactions (~11%) and chemical bonding (10%, Saidin et al., 2019). The last two topics are organic chemistry and the periodic table, each constituting ~5% of AR chemistry higher education studies (Mazzuco et al., 2022). This suggests that AR technology is being used primarily for the 3D visualisation of molecular structures within chemistry education. This is not surprising, as an understanding of the spatial relationships between atoms is essential for understanding the behaviours and properties of a molecule (Johnstone, 1982; Maier and Klinker, 2013). This proportion corroborates the literature which states that the use of 2D images to teach subjects related to 3D molecules limits the ability to understand the main visuospatial elements of macromolecular structures, including the perception of depth, and the sense of scale (Vienne et al., 2020). This reinforces the perception that the understanding of spatial processes and the structures of molecules have been a source of difficulty for students (Fatemah et al., 2020).

From another perspective, it can be assumed that students' difficulties in understanding, interpreting, and translating complex molecular representations (Cai et al., 2014), inherent to chemistry topics such as molecular structures and chemical bonding, may find help with the application of AR within their respective teaching processes. Laboratory practice aims to achieve a similar outcome, through establishing theoretical concepts through concrete experiences. While there are investigations addressing the application of AR in virtual laboratories, these works are limited in number. Domínguez Alfaro et al. (2022) describes the development and testing of a markerless mobile AR application called MAR Lab, using the TrainAR authoring tool. MAR Lab is designed to support students' laboratory skills, with the authors stating that the application can serve as an effective pretraining tool in instances where laboratory access is unavailable. Preliminary results suggest that the application exhibited acceptable usability, but larger cohorts of students are required to understand how students learn with MAR Lab, and if the knowledge learned in the immersive environment can be translated to real-life laboratory situations.

Aw et al. (2020) utilised AR to teach topics pertaining to molecular structures. The development of "Nucleophile's Point of View" (NuPOV) affords students spatial interactions with molecules, supporting self-directed learning. The authors report increased measures of student confidence and good receptivity of the application in understanding nucleophilic reactions. Furthermore, work regarding laboratory learning environments reports the use of AR in Educational Laboratories (ARiEL). ARiEL is used to connect students to information on scientific equipment, afforded by AR technology (An and Holme, 2021). A focus group of students provided usability feedback on the beta version of the application, revealing the ease of use, and students' preference to using AR technology to access information on scientific instruments. Usability evaluations, using the 20-item ASCI (a semantic differential instrument), report reduced anxiety in students, alongside easier and clearer operation of the instrumentation. What these studies demonstrate, in addition to those works outlined in table 2.1, is that the affordances of AR technology, in comparison to classic teaching approaches (Ewais and Troyer, 2019) may enhance important processes such as knowledge retention (Badilla-Quintana et al., 2020; Olim and Nisi, 2020). Furthermore, AR technology may help to reduce extraneous cognitive processing (Buchner, Buntins, and Kerres, 2022), increase motivation (Estudante and Dietrich, 2020; Irwansyah et al., 2018), and enhance spatial skills (Kodiyah et al., 2020). In addition, the simulation of AR laboratory experimentation can lead to substantial benefits in terms of safety, repeatability, and chemical consumption.

## 2.6 Metrics for Learning Success

In a traditional learning environment, students are commonly assessed using a combination of formative and summative approaches. Throughout the academic year, formative assessments are conducted frequently, and are a low-risk, continuous process of observation and informal testing used to monitor students' development (Schildkamp et al., 2020). In contrast, summative assessments formally evaluate students at distinct points in the academic calendar (Dolin et al., 2018). As most research into AR-supported learning environments is cross-sectional in nature, the evaluation of learning success is often determined using multiple choice question (MCQ) instruments, to evaluate short-term improvements in understanding. In addition to conceptual understanding, researchers have also employed the use of metrics such as *perceived usefulness* and *perceived ease of use*, as explained by Davis et al. (1989) in their technology acceptance model (TAM). The TAM is a valid and well-established model that specifies a central theory from the discipline of business informatics. The central questions derived from this model are given in terms of a system's usability, and the aspects of a system that affect a user's attitude and behaviour. In addition, factors such as *perceived learning* and *perceived motivation* are parameters that determine a user's perception, but do not necessarily evaluate learning success.

Perceived usefulness in the context of an AR educational application primarily refers to the extent to which an individual believes they learn the targeted knowledge resulting from interactions with that application. Furthermore, perceived ease of use is based upon the abilities and prior experiences of the user. An individual who is proficient using AR technology will likely experience greater perceived ease of use in relation to a less-experienced user. Noteworthy work regarding metrics for measuring learning success is that reported by Albert and Tullis (2013). Metrics were categorised along the dimensions of 'performance' and 'self-reported', where performance metrics pertain to objective measures, and are always based on the users' behaviour rather than what an individual says. In contrast, self-reported metrics are largely subjective, such as those from Likert scales and semantic differential scales. Performance metrics were typically collected using observation methodologies within specific contexts and settings, whereas self-reported metrics primarily focus on the reliability of a user's opinion. As for data sampling and evaluation in AR studies, three types of evaluation are typically found in the literature, as outlined by Lim et al. (2019):

i. Within-subject evaluations, which are repeated measures on experimental participants that are evaluated on more than one tested item.
ii. Between-subject evaluations, which compares a single evaluation's results between different participants.
iii. A combination of (i) and (ii).

Within-subject evaluations do not generally require a large sample size, but it entails the risk of participant carryover effects. On the other hand, between-subject evaluations typically reduce these risks, but more effort is required to conduct the data collection. Returning to the work of Albert and Tullis (2013), five general performance measures are used in technology-supported environments:

i. **Task success** is used when the researcher is interested in whether a user can complete a task using the technology, based upon a set of criteria. Within an educational environment, this may be based on the degree of completion, a user's experience in constructing solutions to problems, or the quality of the answer provided.

ii. **Time on a task** is helpful when a researcher is concerned about how quickly a user can perform an action using the technology.
iii. **Errors** are based on the number, and nature, of mistakes made by a user while attempting to complete an action.
iv. **Efficiency** is a way of evaluating the amount of effort (cognitive and physical) required to complete an action or task.
v. **Learnability** involves looking at how an efficiency metric changes as a function of time.

In addition, Lim et al. (2019) provide an overview of metrics used in studies that have investigated mobile-based AR learning applications. The authors categorise the collected metrics according to performance versus self-reported metrics, and within-subject vs between-subject evaluation. These include escapism, facilitating conditions, bundled identification, pragmatic quality, stimulation, novelty, price value, and social influence (Lim et al., 2019).

Eye-tracking data has also been identified as an analytical tool that can provide quantitative evidence regarding cognitive processes, with AR-supported learning environments shown to reduce learners' cognitive load (Buchner et al., 2018). However, lower cognitive load does not automatically translate into improved test performance in comparison to control conditions. Furthermore, eye-tracking has been employed to analyse students' use of visual interfaces for interpreting molecular representations (Pienta, 2017); in addition to examining students' interest and engagement when learning collision theory and kinetics (Sweeder et al., 2019), and to investigate the effectiveness of simulations for learning energy concepts in bonding (Vandenplas et al., 2021). Furthermore, Tang and Pienta (2012) found that eye-tracking technology was useful for investigating the effect of question difficulty and cognitive processes. The study reports that unsuccessful students spent more time looking into the solution details, while recording a longer fixation on the questions, as compared to students who obtained higher scores. Lower cognitive load measures did not result in significantly higher test scores or quicker completion times.

Lastly, Williamson et al. (2013) employed eye-tracking to investigate the time students spent examining ball-and-stick representations and electrostatic potential maps. Pupillometric data was used together with gaze data to identify the cognitive load when students answered questions from the Chemical Concepts Inventory (Mulford and Robinson, 2002). Measurements of pupil dilation were found to be promising in revealing important information regarding cognitive processing. Eye-tracking data has shown that AR has the potential to capture learners' key information focus, but that this alone does not necessarily translate into better learning performance.

## 2.7 Perceptions of AR Technology

As indicated by Garzón and Acevedo (2019) a substantial body of literature regarding the use of AR in chemistry higher education corresponds to qualitative reviews. This is evident in several studies (Irwansyah et al, 2018; Lu et al., 2021; Rahmawati, Dianhar, and Arifin, 2021; Wong, Tsang, and Chiu, 2021), which conclude that the inclusion of AR in education is relevant as it improves students' learning achievements and motivation. Exploration of students' perceptions of AR for supporting learning of molecular geometry (Rahmawati, Dianhar, and Arifin, 2021) found that students' learning was enhanced using 3D virtual representational media. Participants of the study commented on the challenge of mentally translating 2D images of molecules into the corresponding 3D representations, in addition to performing rotation operations; and that this difficulty is alleviated through the inclusion of AR. Observational data suggested that the learning conditions were conducive, and that students were enthusiastic about the use of AR-based 3D virtual representation media. Furthermore, students were able to associate spatial information, with the concept of molecular geometry and by extension a molecule's chemical properties, stating that the use of AR was beneficial. Conversely, quantitative data from the study provided evidence of students' capability to translate, and rotate, 3D objects (Rahmawati, Dianhar, and Arifin, 2021). Yet, spatial orientation was the lowest scoring aspect of spatial ability. Thus, the benefit of AR may be most impactful when assisting students with observing the same object from different perspectives.

A further study consisting of 218 participants (Wong, Tsang, and Chiu, 2021) employed marker-based AR cards of organic molecules, allowing students to view and rotate different molecular structures. The authors report that 86% of respondents agreed or strongly agreed that the use of AR technology enhanced their engagement when learning chemistry, with 92% of students agreeing or strongly agreeing that the same AR experience improved their understanding of abstract concepts. Additional statements also referred to pedagogy, with 88% of respondents agreeing or strongly agreeing that the introduction of immersive technologies supported learning through enhancement of the teacher-student interaction. 87% of respondents agreed or strongly agreed that AR technology in the classroom is an effective teaching method for the enhancement of learning. Although this study does not explicitly state whether any learning theories informed the design of the AR experience, data pertaining to system satisfaction was collected. 86% of participants agreed or strongly agreed that the application helped students' understanding of abstract concepts such as hybridization and molecular structure. Moreover 92% of respondents agreed or strongly agreed that they were satisfied with the inclusion of AR within formal chemistry teaching environments. Descriptive statistics around the construct of satisfaction reported by Lu et al. (2021) also show positive perceptions of students engaging with AR technology, especially regarding the design of the UI. The authors provide further descriptive statistics regarding the construct of learning attitude, with reported measures suggesting that students agree that learning about chemistry in AR is a rewarding experience.

Habig (2020) reports the potential benefits of AR representations reported by students for supporting learning of stereochemistry. Participating students saw a potential benefit of AR representations as meaningful supplements for 2D resources. In addition, the authors report students' high interest in learning with AR visualisations. No significant differences were found between male and female responses (Habig, 2020). This is in-line with work published by Nazar et al. (2020), who carried out a system usability study on a developed image-based augmented experience. Participants in this study were recruited from two sources: (i) a Department of Chemistry Education, and (ii) a Faculty of Education and Teacher Training. The authors report that 100% of respondents who implemented AR, as a tool for teaching and studying molecular shapes, found it more interesting and exciting than traditional pedagogical approaches incorporating 2D figures. However, the study does not employ any forms of quantitative instrumentation to examine whether this increased level of interest translates to greater learning gains. All respondents previously stated that they had not used AR as a form of learning media. In addition, Irwansyah et al. (2018) presents a process flow of the stages involved in the development of AR-based learning media. As part of this research, 10 chemistry education students assessed the instructional media. From the scoring criteria, the application scored 89% regarding learning objective relevance, and 75% in overcoming media limitations. These results indicate that AR-based teaching has the potential to enhance the learning of concepts pertaining to molecular geometry.

## 2.8 Advantages and Affordances of AR

Three decades have passed since the term AR was first coined. Yet, despite the promise, AR has only now become widely accessible as a result of the improvements of smartphone technology. The numerous learning opportunities that AR can afford cannot be understated. One essential advantage of AR is its suitability for learners of all ages. Although the focus of this research is chemistry higher education, AR has a distinct benefit in that, unlike iVR, there is not a requirement for expensive hardware, and can be accessed through comparatively low-cost mobile and tablet devices. Not only does this allow for rapid virtual presence through ubiquitous devices, but also the scalability to use synchronously with cohorts common to university-level study. Mobile devices provide many advantages that support AR applications, as they are easy to use, cost-effective, and portable (Cai, Wang, and Chiang, 2014). They provide a high level of social interactivity and independent operability (Hwang et al., 2014); and are useful for outdoor activities (Cai, Wang, and Chiang, 2014), thereby contributing to users' collaboration skills (Bressler and Bodzin, 2013; Yu et al., 2009) and facilitating meaningful learning (Bronack, 2011). Mobile devices afford learner autonomy through the provision of educational resources at any time and location, evident by work reported by Abd Majid and Abd Majid (2018).

Education is one of the most promising application areas for AR, and many researchers have examined its affordances in various learning environments (Akçayır and Akçayır, 2017; Bacca et al., 2014; Chen et al., 2017; Dunleavy and Dede, 2014; Radu, 2014). Given the simultaneous presentation of both physical and virtual elements, AR applications are better grounded in student's regular learning environment. Students can explore theoretical solutions and problem-solving techniques within an immersive hybrid environment which is contextually accurate (Coimbra, Cardoso, and Mateus, 2015, p. 333), thereby facilitating the development of processing skills such as critical thinking, problem-solving, and communicating through interdependent collaborative exercises (Dunleavy and Dede, 2014). Because of the potential of AR to promote critical thinking and problem-solving in a learner-centered environment, the potential applications to learning are wide-spread. AR can encourage students to engage more deeply with a task, resulting in the construction of deep and lasting connections within their knowledge base (Kerawalla, Luckin, Seljeflot, and Woolard, 2006).

Additionally, research has reported that the incorporation of AR motivates students (Silva et al., 2015), encouraging active participation and high interactivity, rather than passivity. In chemistry higher education, reviews have found more affective advantages than cognitive advantages (Mazzuco et al., 2022). The agency of students where they are enacting, developing and determining can encourage deeper understanding (van Haren, 2010). Augmented experiences often require collaboration which provides opportunities for communication with peers, using realistic interactions within a natural interface. Olympiou and Zacharia (2012) found that students in combined environments significantly outperformed those in physical or virtual environments when learning scientific concepts. However, this study overlooked the measurement of affective outcomes. It is important to note that many studies report strong interest and high satisfaction when engaging in AR-supported learning environments (Akçayır et al., 2016; Bacca et al., 2018; Chen et al., 2017). In addition, previous works have reported a medium effect size ($d = 0.72$) for AR, on the impact of learning gains, when using collaborative learning approaches (Garzón and Acevedo, 2019; Ozdemir et al., 2018). Again, this result is unlikely the effect of AR alone, but the combination of different variables that influence AR interventions. However, this informs the importance of AR as a factor for increasing student achievement (Bacca at al., 2018; Ozdemir et al. 2018).

Further, a key pedagogical affordance of AR is the ability to rescale virtual objects, from molecules to planetary bodies. This allows students to better understand, through manipulation, the properties and relationships of objects that would be either too small or too large to examine effectively in their normal day-to-day lives (Johnson et al., 2010). This affordance is well aligned with Malone's (1981) key elements of intrinsic motivation in learning activities; through allowing students an amount of control in their learning environment. This engages students in learning for its own sake, rather than through external regulation. Though other technologies may perform the same function, rescaling in AR systems provides the user a clear representation of spatial and temporal concepts as well as the extra advantage of contextualising the relationship between the virtual object and the real-world environment (Sin and Zaman, 2010). When applied in a situated learning environment, Shelton and Hedley (2004) describe how students were more involved and examined virtual objects more deeply, to obtain the required information.

AR technologies are preferred as an educational tool, not only in chemistry, but in other branches of science (Ozdemir et al., 2018), as the teaching of abstract concepts can be supported by overlaying contextually relevant information. Lin et al.

(2013) state that AR is a supportive instrument for constructing students' own knowledge, in a way that clarifies the relations among theoretical concepts or principles. AR helps to concretise abstract concepts, supports multimodal experiences, and enhances the sense of reality, which in turn is a huge contribution to learning (Ozdemir et al., 2018). These are likely contributing factors to why reported effect sizes for AR learning gains in science, technology, engineering and mathematics (STEM) subjects are higher than those in social science courses.

In particular, AR-related learning outcomes report higher performance while reducing, cognitive load in comparison to other teaching approaches, evident by works such as Bellucci et al. (2018) and Polvi et al. (2018) who report higher performance and lower cognitive load in their AR experimental groups, compared to control conditions. Other cognitive advantages of AR include the enhancement of spatial ability. Hoe et al. (2017) focused on the training of spatial ability using AR, reporting significant effects in favour of the AR condition. Participants reported lower cognitive load, while displaying higher performance, that is, more pronounced spatial ability. However, in these studies, it remains largely unclear why AR should reduce, or otherwise affect cognitive load. One direction of future research would be to compare different AR instructional environments containing materials that incorporate or violate principles from CLT and CTML. To date, the potential of visualisation dominates AR studies in education, and more work is needed to fully understand how to use the characteristics of AR described in Azuma et al. (2001) to also boost AR-enriched learning environments aiming to promote declarative knowledge.

In addition to the learning affordances provided by the conception of AR, another important feature of AR is the provision of collaboration (Billinghurst, Poupyrev, Kato, and May 2000). AR instructional strategies, that are based around collaborative learning, are expected to support the learner-learner interaction for inducing learning motivation (Dalgarno and Lee, 2010). Collaboration was a frequently cited advantage when using AR tools (Bacca et al., 2018). The idea of collaborative learning may originate from the epistemology of social constructivism (Oxford, 1997), which promises to induce learners' motivation and engagement through shared understanding and mutual efforts. Collaboration with access to virtual information allows learners to use non-verbal cues, such as gesture and body language (Billinghurst, Poupyrev, Kato, and May 2000), to enrich their learning interactions and improve communication in addition to verbal actions (Bujak et al., 2013).

## 2.9 The Challenges of Augmented Reality

In addition to the advantages and affordances that AR holds for the education sector, the literature has suggested several directions for future research. Education researchers have published reviews that provide a comprehensive overview of central research topics for the integration of AR into teaching and learning. Yet, many articles in chemistry education do not explain the disadvantages of using AR in teaching. As conveyed by Garzón et al. (2019) and Mazzuco et al. (2022), only a limited number of chemistry education studies (~15%) report the challenges and problems encountered when using AR in an educational setting.

Education researchers who aim to explore AR continually state the valuable contributions that the technology will make to chemistry education. Yet, as outlined by Radu (2012), a comprehensive explication of the educational effects and implications of AR is still missing. As discussed in section 2.6, one challenge relates to the use of performance metrics (Lim et al., 2019). Although AR facilitates data collection for a continuous evaluation of its application in educational settings, new models and methodologies remain to be proposed for using beneficial performance metrics that eliminate the limitations of measures. Becker et al. (2017) suggests the use of personalised student measurements for evaluating teaching and learning experiences that consider the acquisition of skills, competencies, creativity and critical thinking.

Throughout the first decade of the 21st century the relatively high cost of AR technology restricted wider dissemination, until the advent of mobile devices, and the consistent integration of AR onto them (Garzón et al., 2019). Yet, as is the case for immersive technologies, the increased costs and adverse physiological effects, such as dizziness, are still exemplified when compared to desktop monitors. Similarly, a lack of technical standards, due to inconsistent collaboration among companies developing AR technology, and the difficulty in generating meaningful content (Peña-Rios et al. 2013) places undue pressures on educators. The technical challenge of developing AR experiences remains a limitation within chemistry education. The lack of existing AR applications with suitable chemistry content for use in the classroom is an important obstacle. Creating AR apps with such content demands effort and time. Furthermore, the designers of AR technology are usually computer programmers with minimal experience in understanding pedagogical needs. This leads to the creation of AR applications that discerning teachers are unlikely to use.

A major impediment to the effective creation of educational applications is the lack of a conceptual framework (Rasimah et al., 2011). Without frameworks and guidelines to support the development of AR-based educational experiences, the application of technology within the classroom can be superficial and unproductive (Ertmer et al., 2012). Although educators have recognised the benefits of using AR in the classroom, this is vastly impeded by a lack of control over the content in the system, which is a necessity for ensuring mistakes are avoided due to visualisation simplification (Virata and Castro, 2019). These factors are thus to be considered together with the potential gain in learning benefits of immersive technologies to arrive at an optimal decision for the platform to use. Consequently, the need to adapt AR experiences to the requirements of students can be extensive (Wu et al., 2012), restricting the capability to respond to learner differences (Radu, 2012, 2014). Future work should address the affordances of AR that differentiate from other platforms for offering inclusive experiences for individuals with disabilities (Bacca et al., 2018). There is need for improvement of AR technology to accommodate educational content in a simpler way (Sommerauer and Müller 2014).

Negative impacts such as usability difficulties have also been reported. Studies have reported that AR applications can be unintuitive, evident in work by Sanii (2019), who reports that 22% of students find AR too complicated to use. Bacca et al. (2018) reports that few studies have considered the factors of accessibility and usability, yet it is one of the most frequently reported challenges for AR (Akçayır and Akçayır, 2017; Chang et al., 2018; Cheng and Tsai, 2013). The absence of good UI design, alongside the provision of guidance, can make AR technology unnecessarily complex, and thus cognitively demanding. Weak usability also leads to longer activity times, as reported in a case study from Gavish et al. (2015).

Further, technical issues caused by devices that provide AR experiences has been found to lower students' motivation to learn (Wu et al., 2012). Location-based AR has reported issues with tracking, such as static errors that lead to mechanical misalignments or incorrect viewing parameters, and dynamic errors such as delays and motion lags (Cai, Wang, and Chiang, 2014; Cheng and Tsai, 2013). Although technology will continue to advance, and it is expected that these drawbacks will be remedied, future location-based AR should consider these challenges. Physical factors also include the availability (which is lower in comparison to immersive technologies such as iVR) and technical support. Technical issues resulting from AR experiences has been found to lower students' motivation to learn (Wu et al., 2012). Loss of tracking, light dependency, delays in data rendering, battery consumption, and device overheating are commonly reported (Olim and Nisi, 2020). In addition, location-based AR has reported static and dynamic errors such as delays and motion lag (Cai, Wang, and Chiang, 2014; Cheng and Tsai, 2013). As technology continues to advance, the expectation is that these issues will be resolved, yet future AR studies should account for these limitations. Physical factors include device availability, and the provision of technical support.

All the aspects discussed weaken AR technology acceptance, which is crucial to both educators and students to effectively, and fully, exploit the affordances of AR. Clearly, it is necessary that educators need to familiarise themselves with AR technology to effectively utilise it in the classroom, to avoid both considerable obstacles and the increased cognitive processing associated with poor implementation. Therefore, considering teachers' requirements, there will be a need to not only consider the incorporation of interactive strategies to enhance first-hand experience, but also how the AR classroom is designed and evaluated, and the teacher's role within an AR educational setting. Bacca et al. (2018) emphasizes that the conceptualization and construction of tools for teachers to create content requires their involvement in the design of the AR application.

# 3

# AR-Supported Problem-Based Learning Scenarios

In chapter 3, the development and evaluation of my first educational intervention, an immersive technology-supported learning environment based on commercial escape rooms is discussed. An educational escape activity (EEA) was constructed to support students' understanding of stereochemistry, specifically stereoisomerism, structural isomerism, and the rules pertaining to the nomenclature of transition metal complexes (as outlined by IUPAC, 2005). An introduction to Game-Based Learning (GBL) and the pedagogical paradigm of EEAs are outlined in **sections 3.1** and **3.2** respectively, including a critical summary of previous works published in the field of chemistry higher education. Next, details regarding a pilot study conducted on the developed stereochemistry EEA are presented in **section 3.3**. The pilot study served as an opportunity to both evaluate the impact of learning stereochemistry within this active learning environment and observe students' interactions with the immersive virtual reality (iVR) and augmented reality (AR) technology.

The development of an EEA is a continuous, iterative process. Following the pilot study, three design changes were implemented. Firstly, although the pilot study EEA was facilitated synchronously, the transition to remote learning in 2020 restricted the ability to run this activity in-person. Thus, the EEA was migrated to a digital platform as an online experience, overcoming one of the main limitations of the literature – the need for scalability. Secondly, as students were engaged in distance learning throughout academic year 2020/2021, the choice was made to utilise AR technology exclusively within the stereochemistry EEA. At this point, ChemFord was establishing itself within the research project, and focusing development on AR allowed students to engage with my virtual experiences remotely, negating issues of both health and safety, and hardware availability. Lastly, Self-Determination Theory (SDT) was identified as a framework to guide the design of the second, and third iterations of the EEA, to allow insight into the design features that may influence an individual's motivation to engage with the EEA intervention. One current direction of SDT research concerns the promise and problems associated with new technologies for education. One of the great challenges of modern education is that of capturing the attention of students and creating engagement for learning tasks. In response, educators are turning to the attention-grabbing power of games for teaching purposes, using "gamification" strategies to enhance motivation (McKernan et al., 2015). Details of SDT are outlined in **section 3.4**.

The second and third iterations of the EEA were also used as an opportunity to collect quantitative data pertaining to students' learning gains. To complete this, an instrument was developed that could be used to assess students' understanding of inorganic stereochemistry. Details regarding the creation and validation of this instrument are discussed in **section 3.5**. **Section 3.6** presents a second study (covering the second and third iterations of the EEA) carried out in March 2021 and March 2022. An examination of students' performance and motivation measures, in addition to qualitative analysis are discussed. Further, an initial reliability analysis on the created conceptual knowledge instrument, using approaches such as Classical Test Theory (CTT) and Item Response Theory (IRT), is provided. The limitations of the both the pilot and cross-sectional studies are discussed in **section 3.7**, with concluding remarks presented in **section 3.8**.

## 3.1 Game-Based Learning

Students are learners who construct their own understanding and knowledge of the world, based on their own unique experiences. This is one of the core principles of constructivism, most famously voiced by Dewey and Piaget, that learner engagement with the world subsequently constructs meaning through sensory input. Although this predominantly occurs within the mind, there is a necessity to provide learning environments which engage students physically as well as mentally. Consequently, science education researchers are increasingly adopting GBL, the integration of game mechanics into learning experiences, to increase engagement and promote situated experiential learning (Griggs et al., 2019). The denomination originated from Prensky (2003) who popularised the field "Digital Game-Based Learning" (DGBL), the paradigm of adopting digital games for representing and simulating conditions to impart knowledge and nurture social evolution. In contrast to more traditional didactic styles of teaching, GBL can be targeted appropriately to the skill level of individual learning. This often results in effective focus on a task, resulting in deep learning and high levels of satisfaction (Hamari et al., 2016).

The use of play in educational contexts for purposes of learning is not a new concept. Garris et al. (2002) states that games can stimulate motivation, which is one of the fundamental principles of learning. This results in an increase in interest, alongside the promotion of active involvement and students' thinking skills. Further works provide evidence for this, with de Souza Silva et al. (2017) using comparative criteria designed to evaluate students' motivation – specifically the aspects of attention, relevance, confidence and satisfaction – when they use immersive, and non-immersive educational games. Humphrey (2017) uses the term "serious gaming" to describe this link between immersive technology and learning in higher education. In addition, GBL may affect players' values and goals of learning chemistry, which is grounded in motivational theories such as SDT (Ryan and Deci, 2000). However, motivation must be sustained through feedback responses and reflection (Garris et al., 2002). Over two decades later, little is known to what degree design complexity is required for meaningful learning to occur, as GBL is fragmented by learner and design variables. As such, further work is required to ensure designers and educators understand the balance between the integration of game mechanics, and their relation to fulfilling the specified learning outcomes.

The key factors that can impact players' motivation within GBL environments includes adaptive challenge, self-expression, discovery, immersion, collaboration, and low-stake failure (Hamari et al., 2016). GBL also encourages graceful failure, encouraging risk-taking within a safe environment, and the provision of opportunities for self-regulated learning. These aspects all align well with established learning theories such as constructivism. According to Li and Tsai (2013), constructivism is one of the major theoretical foundations employed by GBL researchers within science education, allowing players to set their own challenges and provide feedback to peers through available tools (Plass, Homer, and Kinzer, 2015). Plass et al. (2015) describes the structure of a game adopting a constructivist approach with a simple model (figure 3.1), in which design features are at the centre of the learning experience, permeating how challenge, response, and feedback are designed.

However, one of the main obstacles to the wide uptake of games in learning is a lack of empirical data to support their effective utilisation. Although games are seen as an excellent method of active learning (Dietrich, 2019), only around a third of research studies have found the pedagogy of GBL, within or supported by digital applications, to facilitate students' problem-solving (Li and Tsai, 2013). Yet, chemistry learning involves not only scientific practices, and previous works provide evidence for GBL approaches that support the improvement of spatial cognition, visual attentional processing, and perceptual-motor skills. From the cognitive perspective, GBL may affect processes underlying chemistry learning such as schema construction, grounded in learning theories such as Cognitive Load Theory (CLT) and Cognitive Theory for Multimedia Learning (CTML). Wu et al. (2012) found that GBL tended to yield positive results when learning theories were incorporated into the design, in contrast to simply providing students with a game and expecting increased motivation and knowledge acquisition.

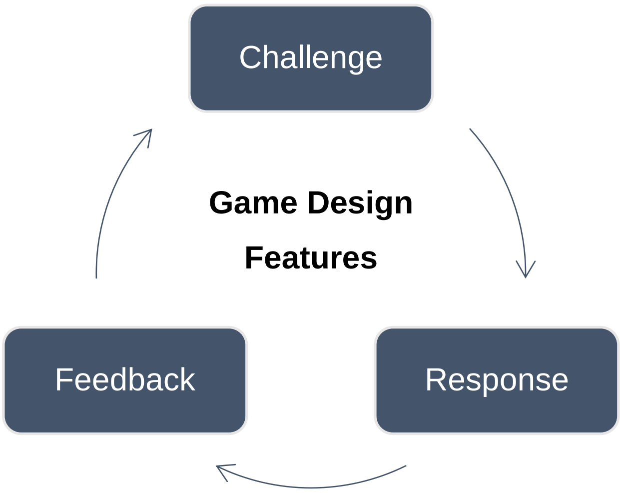



Figure 3.1. Model of GBL as outlined by Plass et al. (2015). A loop is generated when feedback constitutes a new challenge.

However, other studies, such as Wood and Donnelly-Hermosillo (2019), have found no differences in achievement between GBL and non-GBL control conditions, with work by Salomon (1984) reporting lower levels of learning as a response to the introduction of GBL environments. With consideration to CTML, meaningful GBL should occur when a player learns by active processing, and successfully integrates new information with prior knowledge (Moreno and Mayer, 2007). Given that an individual's cognitive capacity is limited, GBL, particularly when multiple representations are involved, can be demanding, and should avoid extraneous processing that does not contribute to learning. Hu et al. (2022) states that chemistry GBL is more effective for cognition and retention in comparison to non-

GBL learning environments. With the development of instructional design, current chemistry GBL environments may be better embedded in learning theories than those reported in previous studies, such as Salomon (1984). Game design and instructional design should aim to optimize cognitive processes and outcomes via the management of essential processing (Mayer and Fiorella, 2014).

## 3.2 Educational Escape Activities

Higher education institutions have experienced a precipitous shift into online learning, with educators facing challenges in maintaining student engagement and motivation. Given the important reciprocal relationship between motivation and learning (Wentzel, 2020), exemplars of multimodal innovations in pedagogical strategy, afforded by advancements in Information and Communication Technology (ICT), have surfaced to improve motivation, whilst supporting students' understanding of chemistry concepts (An and Holme, 2021). One example, built upon the paradigm of GBL is an EEA. An EEA contextualises education content, using game mechanics, into meaningful, collaborative experiences, within a unique learning environment (Tercanli et al., 2021). This sets the groundwork for active learning and social constructivism. Participants accomplish tasks, developed around the subject content, to achieve a team goal within a set time. Hints are provided, when necessary, to ensure that participants remain on track to complete the activity within the allotted time. At the end of the escape activity, participants are led through a debriefing process.

The earliest documented escape room activity was developed by SCRAP (2007), as a single-room activity for teams of 5–6 participants. This model rapidly spread through Asia and Europe, with the World of Escapes directory listing more than 18,000 different escape room activities in more than 45 countries as of June 2022 (Magson and Macpherson, 2022). In educational settings, their potential as a learning activity has inevitably attracted the attention of education researchers. EEAs require participants to collaborate, think critically and laterally, whilst paying great attention to detail (Nicholson and Cable, 2021). While researchers have agreed on the potential and exploration of game-based interventions (Giang et al., 2018), most digital educational games are individual, and do not facilitate collaboration and communication (Dietrich, 2018). Commonly, marks are given based on the final product, not the process. As such, many students would rather work independently, rather than deal with the frustration of teamwork (Williams, 2018). Pedagogically, an EEA attempts to facilitate teamwork. This collaboration, amongst other factors such as the provision of immediate feedback, make them an attractive consideration for in-class learning activities. This is reflected in the literature. EEAs are reported to be positively perceived amongst students (López-Pernas et al., 2021; Oestreich et al., 2021; Vidergor, 2021; Zaug et al., 2021) and have shown great value in terms of engagement (López-Pernas et al., 2021; Oestreich et al., 2021; Vidergor, 2021), motivation (Avargil, Shwartz, and Zemel, 2021; López-Pernas et al., 2021; Oestreich et al., 2021; Ross and de Souza-Daw, 2021), and learner outcomes (Abdollahi et al., 2021; Avargil, Shwartz, and Zemel, 2021; Piñero Charlo, Ortega García, and Román García, 2021). In addition, escape rooms generally have an equal draw from both genders (Clarke et al., 2017). A report of teachers and students engaging in chemistry-based EEAs showed that 96% of respondents considered EEAs to be suitable for developing team building, as well as increasing motivation (96%), and students' communication (95%, Estudante and Dietrich, 2020).

Due to international success as a recreational activity, EEAs are being increasingly adopted by education researchers seeking to increase student motivation and enhance social problem-solving skills (Ang, Ng, and Liew, 2020; Estudante and Dietrich, 2020; Ferreiro-González et al., 2019; Vergne, Simmons, and Bowen, 2019). Previous meta-analysis of research studies demonstrate that greater engagement subsequently results in increased student learning (Freeman et al., 2014). Active learning requires students to participate, unlike previous paradigms such as didactic teaching where responsibility rested with the instructor and the learner played a passive, receptive role. Yet, despite emerging examples of EEAs, few studies have focused on chemistry higher education. Previously reported EEAs have been outlined in table 3.1, alongside further information regarding the incorporated game mechanics and chemistry topics covered. Gamified learning initiatives around the topic of stereochemistry have also been reported previously (Costa, 2007; da Silva Júnior et al., 2017, 2019) but no EEAs around topics of stereochemistry in inorganic chemistry exist in the literature, with minimal studies incorporating the use of immersive technologies. Reviewing educational literature that deals with chemistry-specific EEAs outlines previous examples that serve as:

(i) Instruction to lab techniques (Janonis et al., 2020; Vergne, Simmons and Bowen, 2019; Vergne, Smith, and Bowen, 2020).
(ii) Evaluation of student understanding (Ang, Ng, and Liew, 2020; Clapson et al., 2019; Ferreiro-González et al., 2019).

(iii) Complementary teaching of concepts (Estudante and Dietrich, 2020; Peleg et al., 2019).

Digital variants of the traditional EEA for online learning are the latest development, allowing for the scalability required for implementation in larger educational settings.

The subsequent blending of traditional EEAs and AR technologies is logical, and education researchers are starting to investigate this integration and its influence on learning outcomes (Estudante and Dietrich, 2020; Vicari, 2020; Zeng, He, and Pan, 2020). The comparatively low cost of implementing AR technologies into the classroom on ubiquitous devices provides an opportunity for rapid virtual presence. The vision of the research presented in this chapter draws on the inspiration of using immersive technologies to support this pedagogy. It is important to stress that the EEA is not designed to replace classroom settings, but to add to the overall learning experience.

However, the successful evolution of this teaching approach requires several limitations in the literature to be surmounted. Firstly, the design and preparation of an EEA is very complex and requires time that educators do not have within their working commitments (Cain, 2019; Estudante and Dietrich, 2020; Eukel and Morrell, 2021; Fotaris and Mastoras, 2019; Järveläinen and Paavilainen-Mäntymäki, 2019; Vörös and Sárközi, 2017). A framework for the modular design of an EEA transferrable between different subjects would be significant. Furthermore, larger studies are required to validate the observed results of smaller cross-sectional studies (Clarke et al., 2017; Gómez-Urquiza et al., 2019), alongside the development of EEAs that can be scaled to facilitate larger cohorts (Cain, 2019; Clarke et al, 2017). For cohort sizes common to university settings, EEAs must be scalable without issues such as physical space, resource availability, or time. In addition, previous studies do not currently evaluate how to prevent “free-riding” – team members who do not positively contribute to the tasks within the learning environment. At present, EEAs only evaluate the collected result of whether the team managed to complete the activity within the allotted time. Instead, developers should aim to design EEAs that can evaluate individual competency, to ensure that students’ display the skills and knowledge outlined by the learning objectives. The introduction of roles into the EEA would help to prevent this – the inclusion of tasks that require input from multiple individuals to foster collaboration.

Table 3.1. Previously published EEA studies for chemistry higher education

| Authors (Year) | Purpose of EEA | Participants | Game Mechanics | Success Rate | iVR or AR Used? |
|---|---|---|---|---|---|
| Ferreiro-González et al. (2019) | Escape ClassRoom "CSI: 1.0": An interactive analytical chemistry escape activity for evaluating students at the end of the subject | 43 students | Scenario 1: The Scientific Police Station, Adaptation Zone (Introduce players to the game's plot; Players obtain a secret code to open a locked case and safe-deposit box containing equipment required to complete the activity);<br>Scenario 2: The Crime Scene, Sampling Zone (Evidence collection and application of the sampling procedure);<br>Scenario 3: The Forensic Laboratory, Experimental Analysis, Data Treatment, and Interpretation of the Results (Sample analysis: Toxicological, Chromatography-mass spectrometry, audiovisual, and fingerprinting; Correct analysis provided the final code) | 50% of teams escaped in time | No |
| Vergne et al. (2019) | Laboratory Escape Activity: An escape room activity introduced to foster team building and collaborative learning in a laboratory-experiment setting | Two groups of ~15 students | Obtaining a secret code to access equipment needed to complete the activity; Use of different analytical techniques: UV-Vis, GC-MS, FT-IR; Solving a riddle based on the periodic table | Students finished with times in the range of 35–55 min | No |
| Clapson et al. (2019) | ChemEscape: To engage and have participants enhance and apply discipline specific skills and knowledge | 800 engineering students | The solution to each puzzle is a set of three numbers:<br>Puzzle 1, Thin Layer Chromatography;<br>Puzzle 2, Buoyancy Bottle;<br>Puzzle 3, Density Tower;<br>Puzzle 4: Zinc/Copper Water Cell | ~60% escaped | No |

| Authors (Year) | Purpose of EEA | Participants | Game Mechanics | Success Rate | iVR or AR Used? |
|---|---|---|---|---|---|
| Estudante and Dietrich (2020) | Two scenarios: The LeBlanc process, a physical escape activity (previously published in 2018); The Solway process, an AR escape room application to diffuse AR to a large audience | >70 volunteers | Puzzle 1: an easy puzzle based on the periodic table;<br>Puzzle 2: Finding a molecule for which the CPK colour representation is close to the flag of Belgium;<br>Puzzle 3: Compound identification (writing chemical formulas);<br>Puzzle 4: Obtaining a code from global equation of the Solway process | 95% escaped | AR used |
| Vergne et al. (2020) | ESCAPE from the Chocolate Factory: Online virtual escape room game for an undergraduate chemistry lab class | 8 students | Overall game description using video media; Roaster Room: Determination of molecular weight (with visual aid); Chocolate Processing Room: pH problem with video; Packaging Room: Matching of chromatography terms; Gift Shop: Linear regression problem with video | Students finished with times in the range of 10–20 min | No |
| Ang et al. (2020) | Online virtual escape room to reinforce concepts of chemical bonding | 24 students completed the physical escape room | Puzzle 1: Determination of net dipole moment in molecules;<br>Puzzle 2: Strengths of different chemical bonds and molecular interactions;<br>Puzzle 3: Bond strengths of metallic, covalent, and ionic bonds;<br>Puzzle 4: To teach terminology relevant to chemical bonding; Final puzzle: Determination of the presence and comparison of the strength of hydrogen bonds in molecules | Not explicitly stated | No |

## 3.3 Pilot Study (2020)

To initially understand how different immersive technologies could be effectively embedded into the pedagogical paradigm of an EEA, a pilot study was conducted. A stereochemistry EEA was developed, with virtual elements that could be represented using either AR or iVR technologies. Nine participants, enrolled on module "Bonding, Structure and Periodicity", a compulsory module of inorganic and general chemistry study at the University of East Anglia (UEA), for academic year 2019–2020, were recruited. For students engaging in the EEA employing iVR technology, a health and safety questionnaire (see section 3.3.1 and Appendix A) was distributed to capture measures regarding simulation sickness. In addition, students' perceptions of the learning effectiveness of both the EEA and the utilisation of immersive technologies such as AR and iVR in understanding concepts of stereochemistry are also reported. The pilot study provided an opportunity to not only examine the potential of this active learning environment as an educational tool, but also to observe students' interactions with the immersive technologies.

### 3.3.1 Simulation Sickness

With the development of immersive technologies, motion sickness is no longer confined to travel. As discussed in chapter 2, the utilisation of HMDs within virtual environments as a platform for training, simulation, and entertainment continues to grow. However, one of the major drawbacks is simulation sickness, which can negatively impact user experience, technological acceptance, and safety (Kim et al., 2018). As such, users can rapidly transition from a pleasurable sense of immersion to an aversive sense of discomfort, disorientation and nausea. Simulation sickness is produced by conflicting inputs from visual, vestibular, and somatosensory afferents, which generally carries vestibulo-autonomic responses in humans (Ohyama, 2007). Consequently, virtual environments facilitate the coordination of incoherent visual-vestibular conflict to induce simulation sickness (Akiduki et al., 2003). The symptoms can include headaches, stomach awareness, nausea, vomiting, pallor, sweating, fatigue, drowsiness, and disorientation (Kolasinski, 1995), and are explicitly listed in the "Health and Safety Warnings" accompanying current VR platforms.

Simulation sickness can cause intense discomfort, creating an aversion to further use immersive technologies. Unlike the etiology of motion sickness in vehicles (Rolnick and Lubow, 1991), a defining feature of HMDs is that visual motion inside a virtual environment is controlled by the user (Chen et al. 2012; Stoffregen et al., 2014). In relation to human-computer interaction (HCI) theory, the development and verification of new types of interfaces and interactions within virtual environments attempts to minimise simulation sickness. High-quality tracking systems can minimise the mismatch between a user's visual perception of a virtual environment, and the response of their vestibular system. Furthermore, it has been noted that decreasing the field of view (Fernandes and Feiner, 2016), or lowering the resolution (Carnegie and Rhee, 2015), tends to also decrease simulation sickness. When the field of view is reduced strategically, simulation sickness can be reduced without decreasing a user's subjective level of presence and minimising their awareness of the intervention.

Unfortunately, user experience is not uniformly positive, and controlled research has suggested that simulation sickness is more common in women than among men (Munafo, Diedrick and Stoffregen, 2017). When Koslucher et al. (2015) exposed participants to linear oscillating visual motion stimuli in a moving room, they found that the ratio of motion sickness incidence for women versus men was greater than 4:1. In addition, Read and Bohr (2014) and McConville and Milosevic (2014) exposed standing participants to three-dimensional (3D) stereoscopic films presented via an early version of the Oculus head-mounted display system (the Oculus DK-1). They reported that females were more likely than males to experience discomfort. Interestingly, Zhu et al. (2018) reported no symptoms of simulator sickness when employing the Microsoft HoloLens. Several instruments exist for the measurement of simulation sickness including the Simulation Sickness Questionnaire (SSQ; Balk et al., 2013) and the VR Sickness Questionnaire (VRSQ; Kim et al., 2018). The VRSQ was developed as a more appropriate measure of simulation sickness in virtual environments. The VRSQ was developed as a more appropriate measure of simulation sickness in virtual environments. However, the generalisation of the VRSQ is limited due to the relatively small sample size used (*N* = 24; Kim et al., 2018). As such, for this pilot study, a health screening questionnaire from the School of Psychology at UEA was employed (Appendix A).

### 3.3.2 Experimental Design

The nine recruited participants were randomly assigned to three different experimental groups (figure 3.2) to avoid bias and confounding variables regarding the selection of participants:

**Experimental Group 1** completed the EEA using physical molecular modelling kits. This group was treated as the control throughout the pilot study.

**Experimental Group 2** completed the EEA using ChemFord.,

**Experimental Group 3** completed the EEA using iVR technology. Specifically, participants used Nanome (2022) installed onto an HTC Vive.

Each experimental group participated in only one version of the EEA to eliminate carryover effects. Prior to the EEA, a lecture covering the relevant stereochemistry principles was conducted with the student cohort. Experimental groups 2 and 3 were provided with a short introductory session on how to appropriately operate ChemFord and the HTC Vive respectively to ensure sufficient competency to complete the activity. The health screening questionnaire was circulated to participants in experimental group 3 before, and after, completing the EEA.

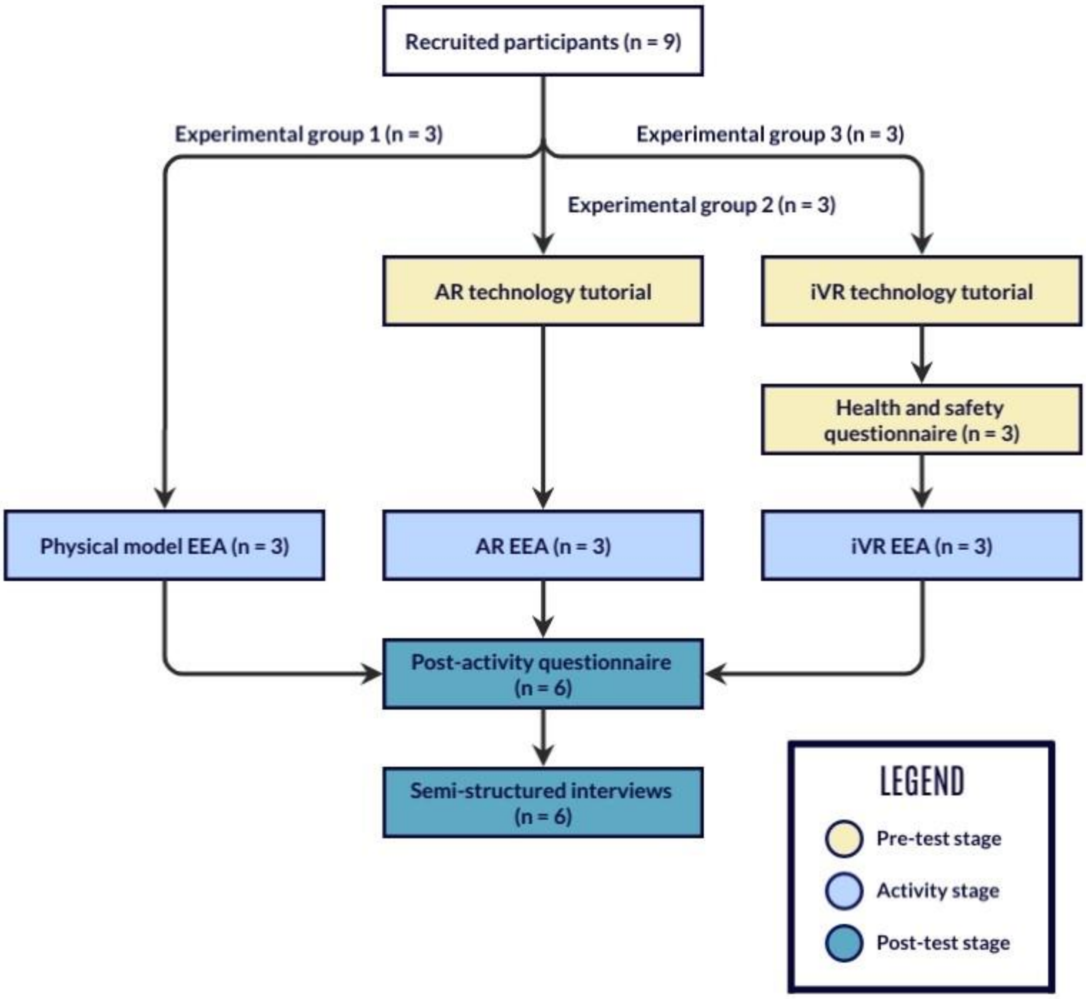



Figure 3.2. The experimental design utilized for the pilot study, including details of participant engagement.

The educational objective of the pilot study was to develop a synchronous EEA to reinforce student's understanding of stereochemistry principles, in addition to developing soft skills such as communication and teamwork. An important initial aspect of EEA design is the creation of an interesting narrative, where the tasks are not only part of the storytelling, but also support participants throughout the learning experience. In a commercial setting, these tasks require no specialised knowledge or skill, with popular task types including "searching for physical objects hidden in a room", "symbol substitution with a key", and "assembly of a physical object" (Nicholson, 2015). For an EEA, the challenge thus evolves to constructing tasks that both fit the chosen narrative, but also sufficiently incorporate the key competencies of the teaching material. The narrative of the activity was based on the recruitment of participants into a fictional secret intelligence organisation. An individual, denoted throughout the activity as *S*, has provided confidential intel highly valued by the organisation. However, this intel has been securely encrypted, and to ensure that the information remains secure, the decryption passkeys that provide access to this intel will be permanently deleted one hour after being first accessed. As the passkeys were chemistry-oriented, the narrative dictated that their skill set was uniquely identified.

The design of the first iteration of the stereochemistry EEA was prompted by the escapED framework (Clarke et al., 2017). The escapED framework was developed to promote a return to inclusive, human-centred interaction within GBL. The overarching pedagogical construct of the initiative is motivated by 'learning by designing', which is a project-based inquiry approach (Clarke et al., 2017), exploiting the characteristics of a non-linear, iterative design process. Comparative work by Neumann et al. (2020), using the escapED framework, looked at synchronous and asynchronous escape activities, with both approaches fostering student engagement and active participation. However, to allow for direct observation, the stereochemistry EEA was designed as a synchronous activity. The escapED program provides a holistic approach to developing learning practice and demonstrates a transition from a technology driven focus to a highly empathetic and person-centred approach. Pre-allocated tips were provided as an attempt to scaffold the progress of experimental groups, to ensure students progressed at a rate sufficient for participation in all aspects of the stereochemistry EEA within the allotted time (60 minutes). Six main areas were considered while developing the activity (figure 3.3).

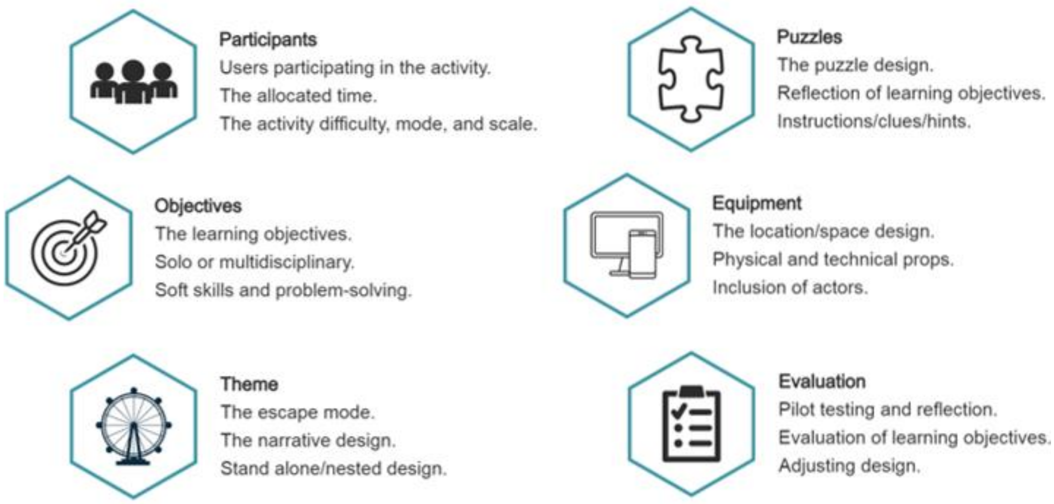



Figure 3.3. The escapED framework adapted from Clarke et al. (2017).

The initial step of the escapED framework is consideration of the participant cohort that would be engaging with the learning experience. The EEA was developed to support students' understanding of stereochemistry principles of transition metal complexes. Designed to be completed in 60 minutes, the activity was constructed to occupy 3 students per instance of the activity. It was important to ensure that participants were challenged, but not to the extent where the activity was impossible to surmount. Hence, tasks with varying levels of difficulty were designed which could be completed in parallel. This allowed students to temporarily step away from a particular task which may be causing frustration, without the worry of losing time or progress. As the EEA progressed to the latter stages, the output of these tasks converged to a single answer that students required to access the final stage (figures 3.4).

Secondly, learner-focused objectives were clarified to facilitate the creation of content. The revised version of Bloom's Taxonomy (1956) was leveraged to produce clear, measurable, and meaningful statements defining the learning objectives. The taxonomy provides a way to organise thinking skills into six different levels: remembering, understanding, applying, analysing, evaluating, and creating (Bloom, Krathwohl, and Masia, 1973). The activity incorporated learning goals that were:

i. Knowledge-related, requiring participants to recognize and identify presented molecular structures.
ii. Analysis- and comprehension-related, requiring participants to inspect and determine different elements of metal complexes.

iii. Evaluation-related, requiring participants to describe and evaluate their progress. The latter prompting higher-order thinking, making participants aware of their successes and mistakes.

Once the targeted principles of stereochemistry had been identified, the tasks were prepared to address the learning objectives. The narrative allowed for the use of a specific style of instruction, as well as the tone used in the scaffolded support.

Paper-based and electronic materials were provided to the experimental groups (figure 3.5). Paper-based clues incorporated the use of confidentially marked files requiring students to not only apply their knowledge of stereochemistry, but also decipher codes to further progress. This is a common component of EEAs (Nicholson, 2015). The electronic tools provided included the ChemFord application (experimental group 2) compiled onto a suite of iPads, and Nanome (2022) installed onto an HTC Vive (experimental group 3). The HTC Vive headset uses six degrees of freedom (incorporation of rotational and directional movement), tracked using the Lighthouse system (Borges et al., 2018). Both ChemFord and Nanome allow students to view and manipulate (rotate, scale, and translate) single, or multiple, virtual representations of transition metal complexes simultaneously. The virtual representations of the transition metal complexes developed for ChemFord were imported into the iVR environment as Protein Databank files (.pdb), allowing them to be rendered within Nanome. To avoid user irritation and ensure that participants' working memory was dedicated to solving the problems presented in the EEA, familiar visual affordances and signifiers were implemented to ensure that ChemFord was intuitive to users. Only essential elements that serve a critical purpose were visible on the UI (Hick, 1952). Experimental group 1 was provided with a physical molecular modelling kit, allowing participants to construct molecules containing either an octahedral, tetrahedral, or square planar central atom. Evaluation of the game experience and learning objectives was embedded into the debriefing session. Specifications of the EEA are outlined below:

**Purpose of the activity:** This activity was designed to evaluate students' understanding of the principles of stereochemistry.

**Goal of the activity:** To successfully solve the task(s) within the allotted time of 60 minutes. The narrative is as follows: *An enigmatic figure known only as "S" has surfaced with information on critical importance. Upon inspection, the information is inaccessible and attempts at brute-force entry have been unsuccessful.*

*Accompanying the information is a series of clues which are believed to hold the key to constructing two passwords, unlocking the contents within. On opening the first of a series of tasks, a countdown will commence, and if all are not solved within 1 h, the information will be lost forever. Will you solve the tasks in time?*

**Activity learning objectives:** Achieving the activity goal supports students' understanding of stereochemistry principles. By the end of the EEA, students will be able to:

i. Differentiate different stereoisomers of transition metal complexes.
ii. Correctly assign the oxidation state of metal atoms bound to ligands within coordination complexes.
iii. Distinguish whether a metal complex is tetrahedral, square planar, or octahedral based on the three-dimensional projections and assign the correct bond angles.
iv. Demonstrate application of the rules of nomenclature to create the name of a transition metal complex (in line with IUPAC recommendations, 2005).

**Tasks within the activity:** The four tasks of the stereochemistry EEA are synopsised in table 3.2.

**Briefing before the activity:** The briefing was used as an opportunity to welcome participants to the activity and present details on how the experience will be structured. The narrative presented to participants directly links into the first task. After this point, the facilitator is no longer directly involved in the experience and only provides support when deemed necessary.

**Debriefing after the activity:** On completion of the activity (or on the expiration of allotted time) the debriefing session commences. This session was treated as an important time of reflection on the learning objectives and to provide feedback on participant performance.

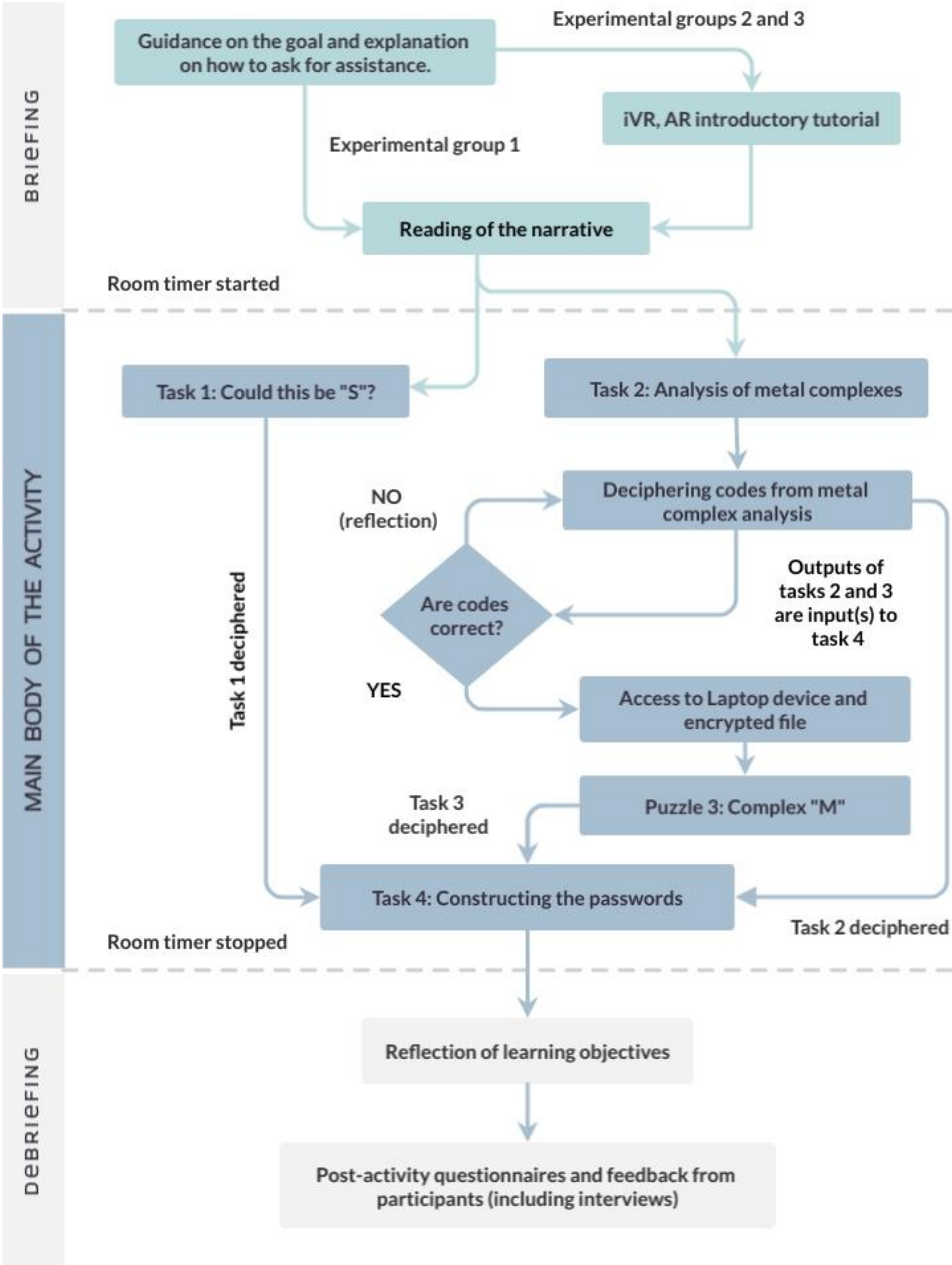



Figure 3.4. The sequence of the EEA employed for the pilot study.

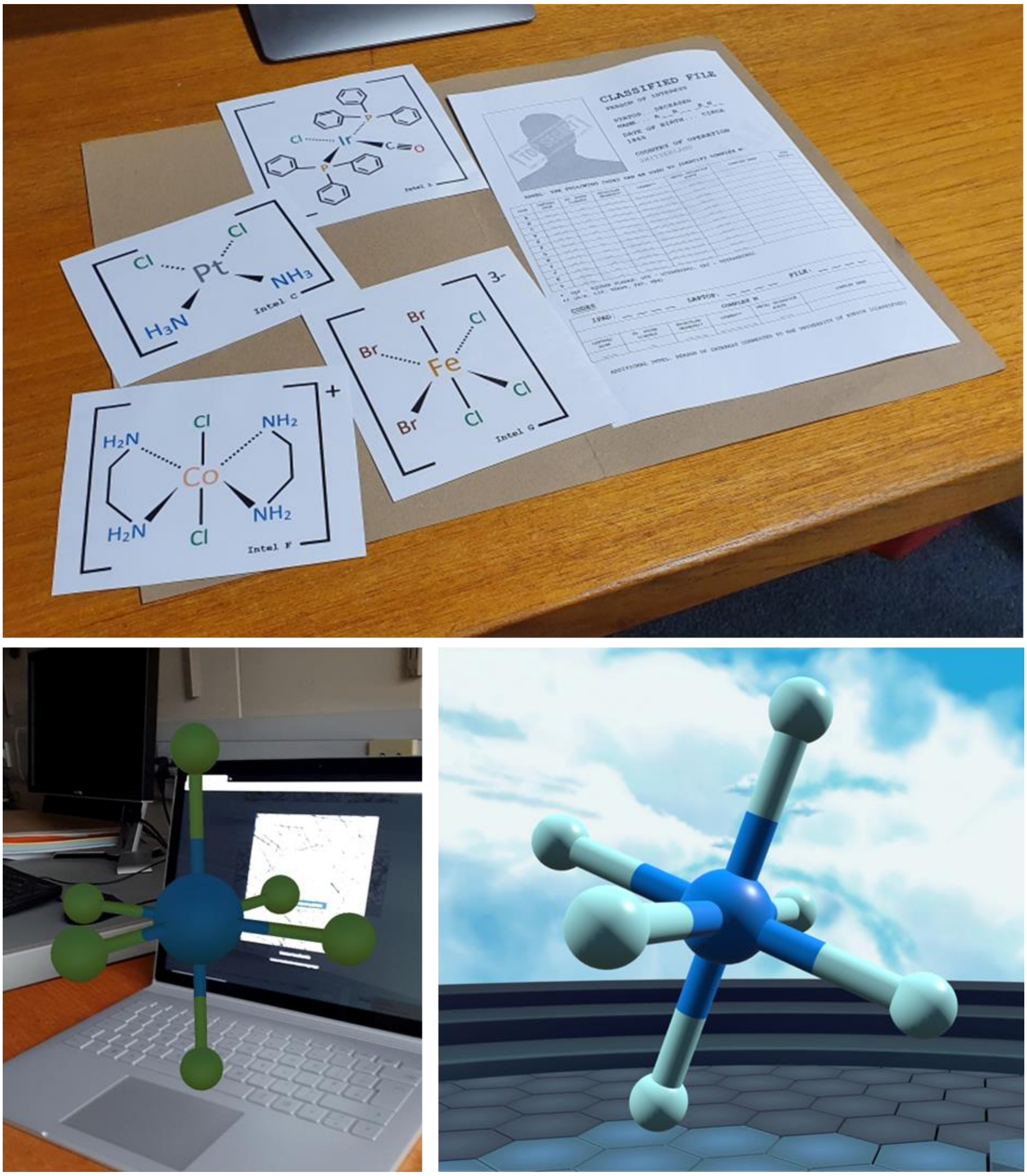



Figure 3.5. Examples of resources used within task 2 (top). A representation of uranium hexafluoride in ChemFord rendered from an image marker located on the laptop display (bottom-left), and the same molecule rendered through Nanome on the HTC Vive (bottom-right).

Table 3.2. An overview of the tasks within the EEA developed for the pilot study.

| Task Number | Type of task | Description of task | Successful outcome of task |
|---|---|---|---|
| 1. Could this be "S"? | Cipher | A simple cipher is used as an introductory task to build student confidence, engage problem-solving skills, and serve as an initial discussion topic. | The cipher is solved, and participants understand how to undertake the remaining tasks. |
| 2. Analysis of metal complexes. | Molecular structure determination | 12 intel image targets are placed throughout the room, each relating to a different metal complex. Participants must extract information pertaining to the central metal atom, bound ligands, isomerism, and molecular geometry. Depending on the group, molecular modelling kits, AR, or iVR is incorporated to assist with identification. | Correctly identifying the metal complexes allows construction of a series of passkeys required to log into a laptop and access password protected files containing the next task. |
| 3. Complex "M". | Confirmation of answers | Students identify the answers from task 2 to construct the passkeys. Access to the laptop device is only possible if the answers to the previous task are correct. Upon accessing the secured file, a 13th complex (denoted complex "M") becomes available. | Participants apply logic similar to that required in task 2 to determine complex "M". |
| 4. Constructing the passwords | Communication, cipher | Information obtained from completing tasks 1−3 is needed to solve the final cipher. If information is missing or incorrect, final passwords cannot be constructed. | The construction of the final passwords is dependent on students effectively communicating with one another. |

### 3.3.3 Developing metal complex virtual objects

The continuous development of the ChemFord application warranted the generation of a series of transition metal complex virtual objects to accompany the tasks within the EEA. To construct these objects, chemical table files (using MDL Molfile format) for each coordination complex were constructed. An MDL Molfile holds information pertaining to:

- The elemental identity of each atom.
- The nature of the bonds within the complex, specifying the connections between atoms and the bond multiplicity.
- The spatial coordinates for each atom.
- Attributes associated with the atoms and bonds (i.e., chirality).
- Attributes associated with the entire structure (i.e., the net charge).

The current *de facto* standard version is Molfile V2000. Figure 3.6 presents the anatomy of a Molfile for uranium hexafluoride ($UF_6$).

```
7  6  0  0  0  0  0  0  0  0999 V2000

 0.0000    0.0000    0.0000 U   0  0  0  0  0  0  0  0  0  0  0  0
-0.2303    0.9810    0.5174 F   0  0  0  0  0  0  0  0  0  0  0  0
 0.2303   -0.9810   -0.5174 F   0  0  0  0  0  0  0  0  0  0  0  0
-0.3990   -0.5664    0.8962 F   0  0  0  0  0  0  0  0  0  0  0  0
 0.3990    0.5664   -0.8962 F   0  0  0  0  0  0  0  0  0  0  0  0
-1.0349    0.0000   -0.4607 F   0  0  0  0  0  0  0  0  0  0  0  0
 1.0349    0.0000    0.4607 F   0  0  0  0  0  0  0  0  0  0  0  0

 1  2  1  0  0  0  0
 1  3  1  0  0  0  0
 1  4  1  0  0  0  0
 1  5  1  0  0  0  0
 1  6  1  0  0  0  0
 1  7  1  0  0  0  0

M  END
```

Figure 3 6. A Molfile for uranium hexafluoride ($UF_6$).

The first line of the Molfile is denoted as the *Counts Block*. The values of 7 and 6 refer to the number of atoms and number of bonds respectively within the structure. Also of note is the third 0 value, which specifies that the structure is achiral. The following seven rows are referred to as the *Atoms Block*. The first three columns outline the cartesian coordinates of each atom, alongside the atom symbol in the fourth column. The following columns (consisting of 0s in this example) are for the specification of attributes such as non-standard isotopes, charge, and valency. The Atom Block is followed by the *Bond Block*. Within this block, the first column states the first atom row number (from the Atoms Block). The value of 1, in this example, refers to the uranium atom in the metal complex. The second column refers to the second atom row number. The value of 2, in this example, refers to the first fluorine atom. The third column denotes how these two atoms are bonded, with values of 1 referring to each dative covalent bond. The fourth column refers to any details regarding bond stereochemistry. Lastly, the *Properties Block* specifies any additional properties not explicitly stated in the previous three blocks. For $UF_6$, there are no additional properties to specify, therefore, the file is terminated using 'M END'.

To translate these properties into a 3D virtual object, they must be imported into a 3D graphics software toolset. For this research project Blender (Foundation, 2022) was chosen, due to my familiarity with the development environment, and its capability to natively export files (using the Blender FBX exporter) to Unity editor, which contains the AR frameworks required to construct ChemFord's augmented experiences. To import the Molfile contents into the Blender environment, the Molfile was first converted into the Protein Data Bank (.PDB) textual file format using Open Babel v2.3.1 (2022). The output when importing the $UF_6$ file into Blender is shown in figure 3.7.

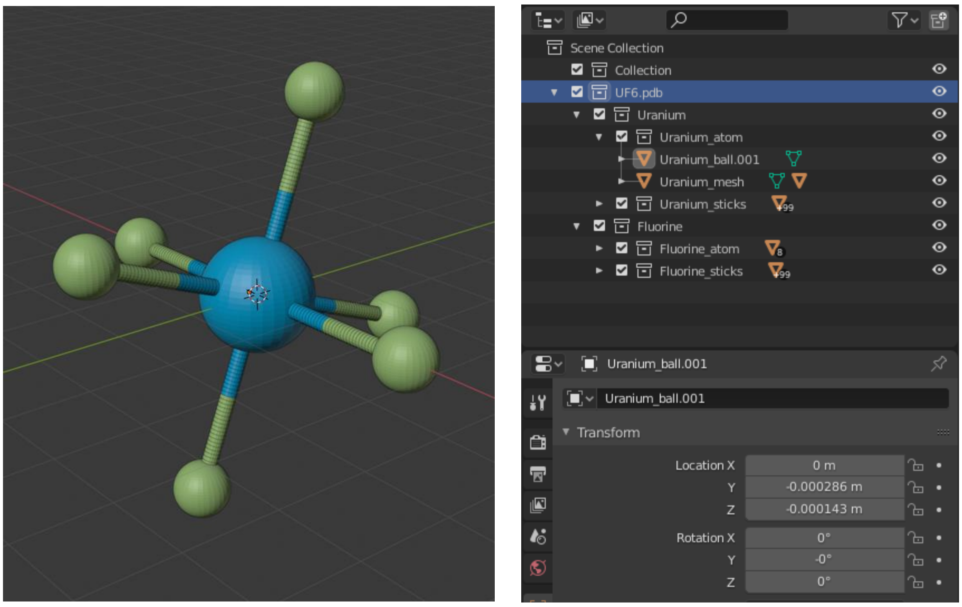



Figure 3.7. Representation of the $UF_6$ PDB file in Blender (left), and the scene collection (right).

The scene collection displays four different independent parent objects, referring to the uranium atoms and sticks, in addition to the fluorine atom and sticks, containing numerous child objects. To consolidate these objects as one virtual object within the augmented environment, the vertices of all child objects were instanced, and then joined, followed by decimation (factor = 0.75) to minimise the vertex/face count. The result is a $UF_6$ 3D model which can be saved in Filmbox format (.FBX) and exported into the Unity editor (figure 3.8).

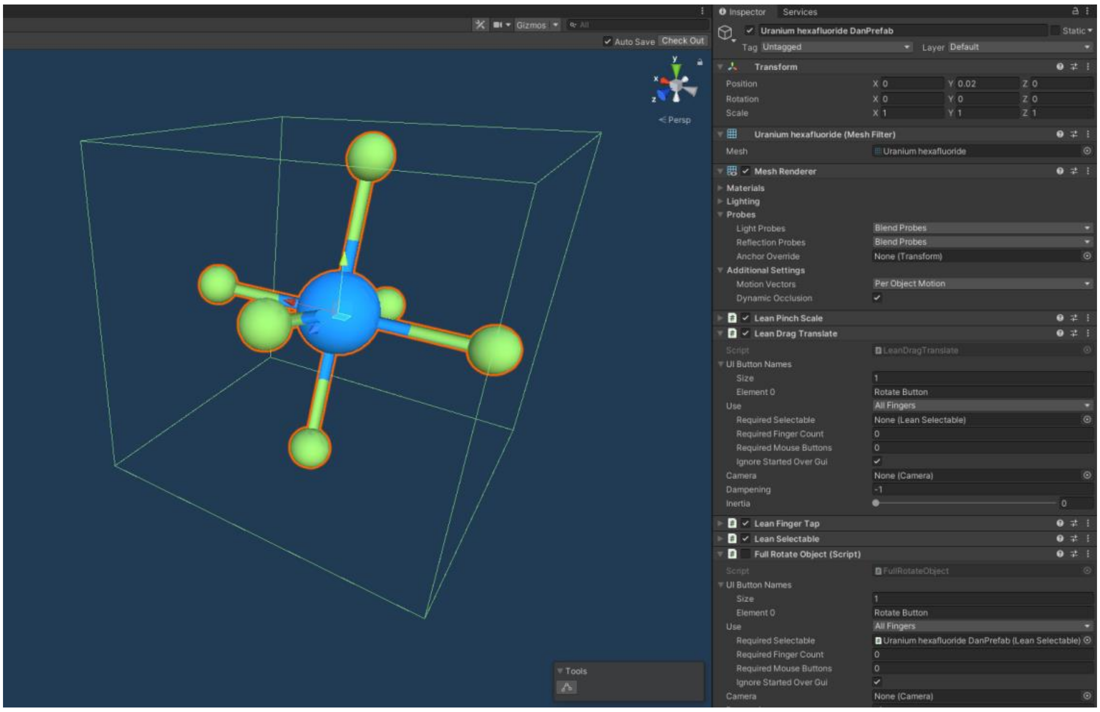



Figure 3 8. $UF_6$ FBX object imported into the unity editor, showing some object components.

Within the unity editor, several components can be attached to the $UF_6$ object to expand its functionality (found in the inspector on the right-hand side). Components common to most objects in ChemFord include colliders for detecting raytracing, and scripts allowing for virtual objects to be independently rotated, translated, and scaled.

A further component inherent to all objects contains information pertaining to the object's image target found within ChemFord's database. When the image target is detected, this component spawns the virtual object, and tracks it within the environment. For the stereochemistry EEA pilot study, the image targets were 13 pieces of intel found within the learning environment, each representing a different transition metal complex.

### 3.3.4 Questionnaires and Interviews

To explore the students' perceptions of the stereochemistry EEA, a questionnaire was constructed, and distributed to participants following the activity. The questionnaire was voluntary, and available to each experimental group for up to one week following the intervention. Eight different constructs were adapted from literature to examine the theoretical, and practical, underpinnings of the activity, in addition to the AR and iVR tools implemented (table 3.3). Students were required to rate their experience using a 5-point Likert scale, ranging from strongly disagree to strongly agree (figure 3.8). Semi-structured interviews were also employed. The interview schedule introduced two further constructs:

i. Perceived usefulness (Davis, 1989).
ii. Representational fidelity (Dalgarno, Hedberg, and Harper, 2002).

Details regarding the interview schedule can be found in Appendix B. Prior works have shown the derived constructs to influence learning effectiveness when employing GBL pedagogical strategies.

Table 3.3. Questions constituting the post-activity questionnaire.

| Construct | Item | Question source(s) |
|---|---|---|
| Control and active learning | This type of learning experience helps to get myself engaged in the learning activity. | Adapted from Lee, Wong, and Fung (2010) |
| Cognitive benefits | This type of learning experience makes the comprehension of material easier. | Adapted from Antonietti et al. (2000) |
| Immediacy of control | The ability to change the view position of the 3D objects allows me to learn better. | Dalgarno et al. (2002) |
| | The ability to manipulate the objects (pick up, change size) makes the learning experience more motivating and interesting. | |
| Motivation | Learning using this tool was fun. | McAuley et al. (1989) |
| | After trying this type of learning tool for a while, I felt pretty competent. | |
| | This type of learning experience did not hold my attention. | |
| Perceived ease of use | Overall, I think that this type of learning tool is easy to use. | Davis (1989) |
| Perceived learning effectiveness | I learned a lot of factual information on this topic. | Benbunan-Fich and Hiltz (2003); Marks, Sibley, and Arbaugh (2005); Martens, Bastianens, and Kisrcher (2007) |
| | I was able to summarize and conclude what I learned. | |
| | I was interested and stimulated to learn more. | |
| | The learning activities were meaningful. | |
| Reflective thinking | I was able to link new knowledge with my previous knowledge and experience. | Maor and Fraser (2005) |
| Satisfaction | I was satisfied with this type of computer-based learning experience. | Chou and Liu (2005) |
| | I was satisfied with the teaching methods in this type of computer-based learning experience. | |

### 3.3.5 Results

Three experimental groups, consisting of three students each (n = 9) attempted the stereochemistry EEA. Of these, experimental groups 1 and 3 successfully completed the activity in 57 minutes and 46 minutes respectively. Experimental 2 failed to complete the activity within the allotted time. Six students completed the post-activity questionnaire and agreed to an interview. The respondents were from experimental groups 2 and 3. Students had a very positive perception of the EEA. All respondents stated that they either agreed or strongly agreed that the learning experience was engaging, and all strongly agreed that they were interested and stimulated to learn more.

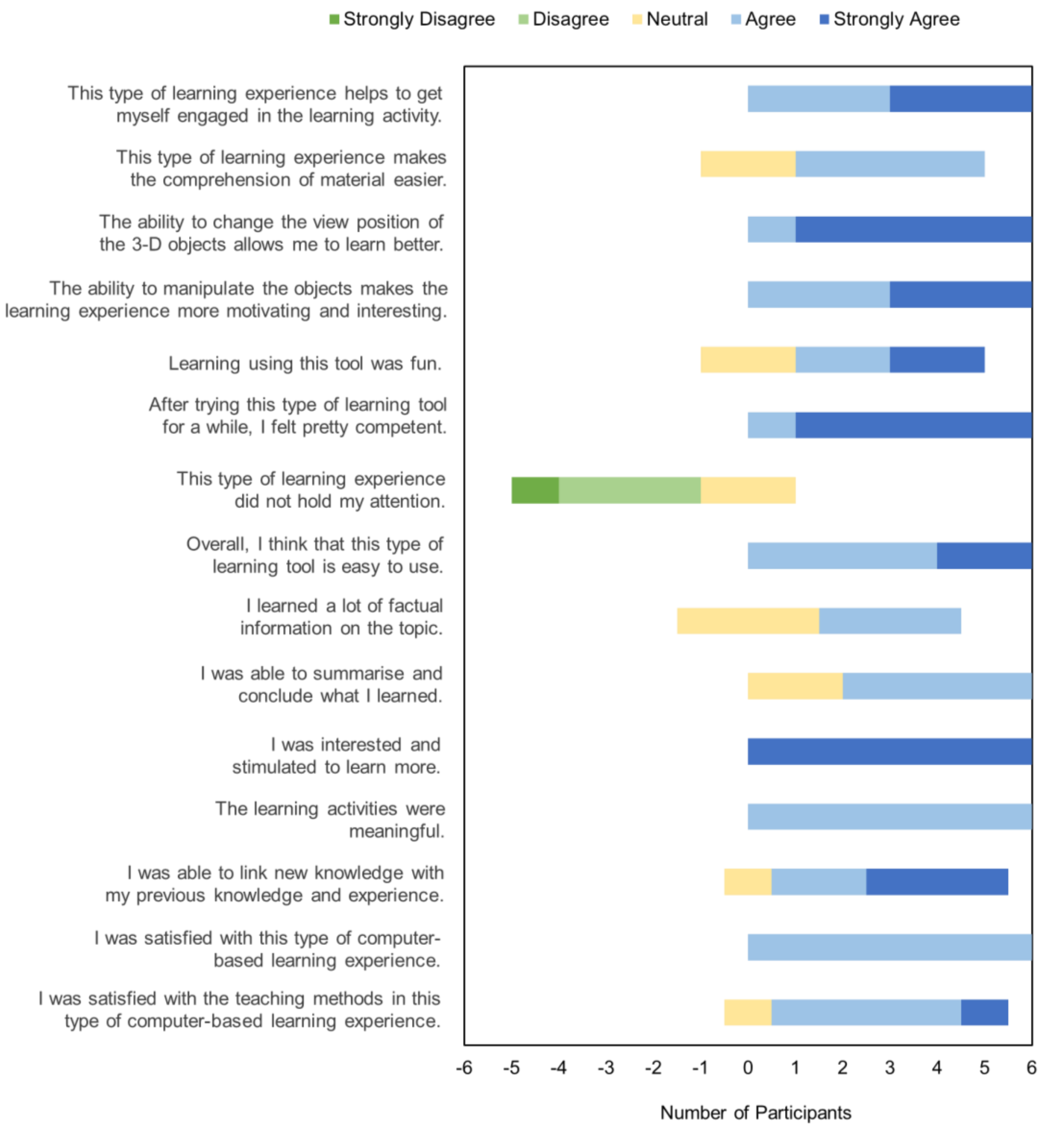



Figure 3.5. Reported measures from the post-activity questionnaire regarding the pilot EEA.

Furthermore, students agreed that the activities were meaningful in achieving the intended learning outcomes. The pilot EEA was designed with a focus on content interaction through engagement with multimedia. A key pedagogical affordance of ChemFord and iVR is the ability to rescale virtual objects, allowing students a better understanding through manipulation of objects that would otherwise be imperceptible through interaction with the physical world. The construction of task 2 placed both ChemFord and the iVR technology as core tools, with this key affordance central to its design. Participants either agreed or strongly agreed that, after experiencing ChemFord and the iVR technology, provided throughout the activity, they felt competent.

In addition, participants either agreed or strongly agreed that the learning tools were easy to use. Interview respondents indicated that the ChemFord application was 'intuitive' with no observable frustration from students using ChemFord. Participants agreed that they were satisfied with this type of computer-based learning experience. It was further observed, throughout the activities, that participants utilising ChemFord and Nanome engaged in deeper discussions regarding the properties of each transition metal complex with their peers. Discussions assisted participants as they used the technologies to dissect each 3D virtual complex and evaluated different structural properties such as the adopted geometry and exhibited isomerism. Throughout, evidence of intellectual quality was apparent as participants constructed and validated solutions to each problem based on substantive communication with their group members. Students believed that the immediacy of control positively impacted their learning. The ability to manipulate and change the view position of the 3D objects positively affected the learning experience to make it both motivating and interesting.

Most participants agreed that the learning experience made the comprehension of material easier, but also expressed neutral responses when asked if the activity provided opportunities for individuals to learn a lot of information. Participants' responses suggest that the EEA primarily functions as an opportunity to consolidate and test understanding. One of the key questions following the pilot study was how the EEA can be successfully implemented within the teaching and learning process, and how immersive technology can support this initiative. Furthermore, even though it was not a primary research area for this pilot study, the opportunity was taken to gather data regarding simulation sickness, as studies increasingly outline reports of health concerns among users of HMDs (Kim et al., 2018). All participants from experimental group 3 completed the health survey. No "severe" symptoms were reported and of those symptoms reported as “slight” or “moderate”, post-activity symptoms had not increased in severity (table 3.4).

Table 3.4. Results of health screening questionnaire

| **Symptom** | **Severity of symptom** | |
|---|---|---|
| | Pre-EEA | Post-EEA |
| General Discomfort | None | None |
| Fatigue | Moderate | Moderate |
| Headache | Slight | Slight |
| Eye Strain | None | None |
| Difficulty Focusing | None | None |
| Salivation Increasing | None | None |
| Sweating | Slight | None |
| Nausea | None | None |
| Difficulty Concentrating | Slight | Slight |
| Fullness of the Head | None | None |
| Blurred Vision | None | None |
| Dizziness with open eyes | None | None |
| Dizziness with closed eyes | None | None |
| Vertigo* | None | None |
| Stomach Awareness** | None | None |
| Burping | None | None |

* Vertigo is experienced as loss of orientation with respect to vertical upright.
** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea.

### 3.3.6 Student Perceptions of the pilot EEA

Qualitative analysis of the participant interviews was completed through latent thematic analysis using the approach of Braun and Clarke (2006). All collected data was transcribed verbatim, coded, and subsequently grouped by themes. The initial broad themes were constructed based on the frequency and similarity of responses, which were then collapsed into core themes by eliminating redundancies and merging closely related major themes. Frequencies were used to highlight important areas for theme development. The process attempts to go beyond the semantic content of the data, and to identify underpinning theoretical ideas. The data suggested several characteristics that appeared to influence students' perceptions of the effectiveness of both the EEA as a learning environment, and the immersive technologies integrated. Pilot study interviewees (PSI), for purposes of pseudonymisation, are represented by a number. The first theme denoted **cognitive benefits** related to the impact that ChemFord and the iVR technology had on assisting students in appreciating the 3D structures represented by the 2D isometric image targets provided. The immersive technologies assisted mental visualisation of the 3D metal complexes, reducing students' difficulty in dealing with abstract concepts such as isomerism and bond angle identification. This was apparent throughout the activity, where participants using the immersive tools exhibited greater confidence and competency when discussing the tasks with their respective team members.

*"...I do struggle to visualize things but if I do go through questions, with time, I can usually get the answer. However, with this, I can visualize them better..."* (PSI 5).

*"Honestly, it was easier, and I liked how interactive it was. I feel like I am more likely to get the wrong structure when using the molecular models... …Being able to move it and see the isomer itself is useful..."* (PSI 2).

Overall, students expressed more positive views towards the AR and iVR tools utilised in EEAs completed by experimental groups 2 and 3, in comparison to experimental group 1, who utilised traditional molecular modelling kits. It was interesting to note that participants also commented that ChemFord would be a welcome addition to their synchronous sessions. ChemFord not only aroused interest and curiosity, but also encouraged active learning through interaction. Due to its adaptability, not only can it be easily upscaled, but also made available for use outside of formal learning environments. The second theme of this thematic analysis is **perceived learning effectiveness.** A minority of my participants expressed low levels of interest regarding the topic of stereochemistry prior to partaking in the EEA but articulated that they understood the importance of the topic. Some participants felt that they possessed low visualisation skills, which was a central source of frustration.

*"Personally, not that interested, I can see how it's useful—I'm not very good at visualising..."* (PSI 6)

However, those who reported lower levels of interest regarding the subject of stereochemistry commented that they experienced greater levels of engagement and interest following completion of the pilot EEA. This resulted in a perceived benefit to learning. One of the main purposes of the pilot EEA was to motivate students, to involve learners who are more reserved in the learning environment. In contrast, participants also perceived a disadvantage to the EEA regarding the initial structure of the activity. It was acknowledged that after the initial briefing, students were left to investigate and analyse the first task. Many stated that the uncertainty of how to progress through the initial stage of the activity induced nervousness. All respondents exclaimed that within the first 5-10 minutes, this feeling had subdued. Most participants also stated that they found the activity a better method to consolidate prior knowledge. Next, the theme of **perceived usefulness** was identified**.** Though participant responses suggested a clear link between initial student nervousness and the structure of the EEA, the experience was viewed as challenging and fun.

*“I thought it was fun, at first, I was nervous, because I didn’t know what I was doing in the beginning...”* (PSI 2).

All participants expressed that they would like to see both ChemFord and iVR tool, plus EEAs, implemented throughout further areas of the chemistry undergraduate syllabus. Most students also commented that the methods employed may help engage others when trying to discuss chemistry outside of the classroom. Naturally, seldom can students incorporate the discussion of concepts like stereochemistry into spontaneous conversation with others outside of formal education.

*“If I wanted to get someone to be more interested in chemistry, like my family; as in trying to talk about chemistry, and I did this activity, they might be more interested…”* (PSI 1)

The fourth theme identified was **representational fidelity.** Many participants expressed preference when using ChemFord as they perceived clear-cut advantages when compared to the alternate methods. Speed, the convenience of generating molecular structures through scanning available image targets, and the ability to manipulate 3D virtual objects were major incentives for using AR technology (figure 3.9).

*“I think it makes it quicker for me to visualize it, with just a picture, it takes me a few minutes to be able to visualize it. It’s quicker in my mind...”* (PSI 3).

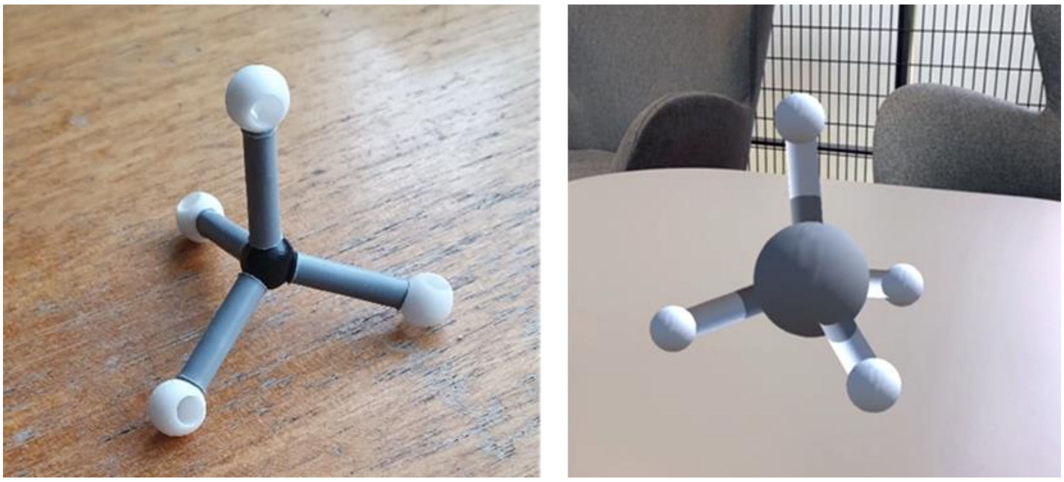


Figure 3.9. Physical model of methane (left); methane as virtually represented in ChemFord (right).

Participants considered these advantages paramount in their own ability to arrive at the correct answer quicker than if using physical molecular modelling kits. No preference was reported in terms of visualisation between ChemFord and Nanome on the HTC Vive.

*"I think I would prefer to use the iPads over physical models... …It takes less time... …Being able to rotate them was great..."* (PSI 2).

The last theme generated in this thematic analysis was **satisfaction.** Participant responses reveal that GBL actions such as the EEA enhanced engagement and improved motivation, group work, communication, and commitment to the learning tasks within the activity. The introduction of the element of time pressure was seen to enhance motivation and competitiveness.

*"Yes, I thought it was good, and that I worked better under the stress. I also felt more active because I was moving around when doing the work…"* (PSI 1).

Students expressed that the active nature of the learning environment also encouraged collaboration, which was apparent through observation. The teams worked well when there was a common context for communication.

*"It's quite cool. It is more of a group activity than just doing questions. It's easier to work together than just having a piece of paper..."* (PSI 5).

To conclude the semi-structured interviews, participants were asked to recount the worst aspects of the EEA. All participants answered that they did not think there was anything they disliked, and all would recommend the activity to other students.

### 3.3.7 Discussion

There is a need to create innovative teaching strategies that can be merged with immersive technologies (Blessinger and Wankel, 2012). The development of EEAs embedding elements of immersive technologies can merge the educational qualities of games, the teamwork and problem-solving skills associated with commercial escape rooms, and attractive technologies to generate effective learning activities that are appealing to students. The pedagogical approach of GBL was employed to create, implement, and evaluate an EEA incorporating a serious gaming strategy. The activity adequately covers the content seen in the classroom and can be used as a complementary tool that helps students reinforce their understanding of stereochemistry principles.

Students' perceptions to the developed activity were captured through interview and evaluated using methods of qualitative analysis. Interview responses were very positive, and the initiative was very successful at capturing students' interest, with participants strongly agreeing that they were interested and stimulated to learn more. These findings are in line with previous studies where educators have incorporated games into the teaching process to aid students with reviewing and reinforcing stereochemistry topics (da Silva Júnior et al., 2019). Similarly, from survey feedback and informal observation, participants were shown to be highly engaged and active throughout the learning experience. Extrinsic motivational factors such as time constraints and competition were mentioned by interview respondents.

Participants expressed initial nervousness when participating due to the absence of a clear path of progression, a property which is commonplace with commercial escape rooms. Not only does this result in the loss of potentially significant amounts of time initially, but it may also lead to demotivation and frustration in students. However, as the activity progressed from the briefing to the first task phase, it was observed that students became more confident and comfortable with the activity once they had established an understanding of what was required to progress - whether this was the discovery of an important piece of information, or resource. To aid the transition from the briefing phase to the initial task phase, the briefing session should address, at minimum:

i. The narrative, if incorporated into the activity.
ii. Guidance on the goal and how to ask for assistance when required.
iii. An indication of which activity mechanics the students should initially focus on.

Design choices should create subtly inform students of the next action they should take. Another suggestion is to make the first task very easy to further facilitate participants getting started with the activity. Students reported a perceived benefit of interacting with 3D virtual representations, and strongly agreed that the AR and iVR technologies supported them when visualising abstract concepts. However, due to the lack of quantitative data within this study, it was not possible to establish sound statistical evidence of improved performance. The EEA grants participants complete freedom, and the social nature of the game allows students to learn in a cooperative environment. How the observed group dynamic within the environment influenced the outcome performance was noteworthy. Experimental group 3, the best-performing group in terms of time, were very vocal among team members with their discoveries and progress. The EEA stimulated students to discover as a team, providing the opportunity to develop adaptive and responsive skills expected of each participant. Experimental group 2 failed to complete the activity within the allotted time, not due to misunderstanding the conceptual nature of the chemistry topics covered (where they scored very highly), but due to an inability to overcome the game mechanics. This highlights a need to ensure the integration of meaningful game mechanics relevant to pedagogical objectives, whilst avoiding the superficial (Arnab et al., 2015).

The process of developing the resources required to facilitate the EEA raised key discussion points. Generating EEAs can require potentially expensive resources, and any permanent physical installation is likely to be unsustainable. Therefore, EEAs should be portable and sustainable. The low financial cost of developing the EEA used in this study is a hugely positive aspect when compared to other GBL activities. Construction of the paper-based resources utilized materials that are commonly available in practical classrooms. Lamination of paper-based resources to extend their reusability was the highest direct financial cost. ChemFord was constructed, in its entirety, using free, available software, and can be downloaded directly onto students' personal devices. However, it is recognized that not all educational institutions will have iVR hardware, such as the HTC Vive readily available.

Refinement of the EEA focused on the capability to support larger cohorts of students simultaneously. This will be a requirement for the incorporation of any GBL activity into mainstream teaching. Most reported studies in the literature are composed of participant groups of 3–7 students per session (Fotaris and Mastoras, 2019). However, for larger groups of students, which is common in a university setting, facilitators must expend considerable effort and time over multiple sessions. The incorporation of augmented technologies and online collaborative spaces is a potential solution to this challenge. The activity is easily portable, and the entire contents could be easily carried in one box by a single facilitator. For this study, multiple classroom locations were employed to ensure that different locations could house the activity. Although experimental groups in this study were small (with only one facilitator present) it would need to be seen whether further facilitators would be required upon scaling of the activity. In this instance, the event was manageable. Although the briefing and core activity could be completed in the 60-minute session, additional time was required to complete the debriefing session. To compensate for the extra time required, future iterations of the escape activity will modify the task mechanics to ensure that the briefing, core activity, and debriefing segments can all be completed within the allotted session time.

During construction of the tasks, it is important for EEA designers to understand how the difficulty of each task should be set to reflect both the task mechanic and the subject material. A task that is too difficult will result in frustration, anxiety, and demotivation, and may even result in students being unable to complete the activity, whereas tasks that are too easy will not provide students with sufficient satisfaction. Prior works have reported the percentage of students who successfully completed the researchers escape room intervention but fail to provide information about those students who were unable to complete the activity. Such data is essential to enable the evaluation of EEAs to enable improvements in subsequent iterations. Pilot testing was essential in the iterative construction of the activity and revealed the requirement of scaffolding to provide guidance to participants. The management of scaffolded guidance is important for the success of commercial escape rooms, but prior works have not presented significant research into the incorporation of such management systems. Common methods include providing guidance on demand when asked by the players or providing guidance when considered necessary by the facilitator. Previous studies have implemented guidance into an EEA that required students to pass a small quiz to earn the right to get help from the instructors (López-Pernas et al., 2019). In an educational setting, guidance steers participants to complete the activity within the allotted time. Although not a primary research goal of this study, further work into the development of innovative guidance management systems for GBL educational settings is required. For successful implementation of EEAs, three considerations, formulated as a result of the pilot study, that require attention are:

1. How is the EEA positioned within the holistic teaching and learning process and what are its requirements?
2. How does the EEA session evaluate individual participants to ensure knowledge and skill competency has been achieved?
3. How can an EEA incorporating AR/iVR tools be upscaled to accommodate larger groups of concurrent players?

## 3.4 Self-Determination Theory

Many educators are concerned with motivational research (Huang et al., 2018; Liu et al., 2013; Reeve, 2012; Ryan and Deci, 2020). The interplay between the extrinsic influences acting on an individual, and their intrinsic motives is central to SDT, a framework for understanding factors that affect an individual's inherent motivation (Deci and Ryan, 2014). SDT is an organismic dialectical approach, meaning that people are considered as active organisms, with evolved tendencies towards growing, mastering ambient challenges, and integrating new experiences into a coherent sense of self. These natural tendencies do not operate automatically, but instead require ongoing social supports. Relevant to this trend is a substantial body of SDT research demonstrating how features of games that satisfy autonomy, competence, and relatedness needs account for the motivational draw of successful video games (Ryan and Rigby, 2019). Students' and teachers' motivation to use technology as a tool for learning will become an even more active area of research (Peters, Calvo, and Ryan, 2018; Sørebø, Halvari, Gulli, and Kristiansen, 2009). As such, the second iteration of the EEA was constructed using the framework of SDT, with consideration towards the affordances of AR technology, to motivate engagement and learning.

The notion of intrinsic motivation is extremely relevant to educational settings and has been shown to be consistently associated with higher performance (Taylor et al., 2014). These intrinsic motivations are not necessarily externally rewarded or supported, but nonetheless they can sustain passions, creativity, and sustained efforts. The basic premise of SDT is that it is not the amount of motivation, but the nature of distinct motivational types that holds the most predictive and explanatory power as to how people behave (Deci and Ryan 2008). Because intrinsic motivation is fully autonomous, it is seen as the ideal motivational type to drive actions (Vansteenkiste et al. 2009). On the other hand, extrinsic motivation is derived from extrinsic regulations that are not related to the activity concerned (Otis et al. 2005; Reeve, Deci, and Ryan 2004; Vansteenkiste et al. 2009). These regulations are external cues that form an outside pressure controlling someone to conduct a desired behaviour. SDT articulates:

i. A meta-theory for framing motivational studies.
ii. A formal theory that defines intrinsic and varied extrinsic sources of motivation.
iii. A description of the respective roles of intrinsic and extrinsic motivation in cognitive and social development, and in individual differences.

Perhaps more importantly, SDT propositions also focus on how social and cultural factors facilitate or undermine an individual’s sense of well-being and the quality of their performance. Expanding further, those who experience pressure from external regulations to conduct a desired behaviour, who are extrinsically motivated, are very likely to feel an innate need to internalise these regulations (Organismic Integration Theory, table 3.5). The more successful the process of internalization, the more these sub-optimal extrinsic regulations echo the characteristics of intrinsic motivation. Hence, it is argued that the addition of points and leader boards to a system reduces gamification to a meaningless ‘pointification’ with little to aversive effects (Roy and Zaman, 2017). SDT assumes that humans are inherently prone toward psychological growth and integration, and thus toward learning, mastery, and connection with others (Ryan et al., 2019).

Table 3.5. Intrinsic and extrinsic motivations, alongside associated processes, and the perceived locus of causality. Adapted from (Ryan and Deci, 2020).

| **Amotivation** | **Extrinsic motivation** | | | | **Intrinsic motivation** |
|---|---|---|---|---|---|
| | External regulation | Introjection | Identification | Integration | |
| | Internalisation → | | | | |
| Lack of perceived competence.<br>Lack of value or nonrelevance. | External rewards or punishments.<br>Compliance.<br>Reactance. | Focus on approval from self and other. | Personal importance.<br>Conscious valuing of activity.<br>Self-endorsement of goals. | Congruence.<br>Synthesis and consistency of identifications. | Interest.<br>Enjoyment.<br>Inherent satisfaction. |
| Impersonal | External | Somewhat external | Somewhat internal | Internal | Internal |

Amotivation is characterised by the absence of both intrinsic and extrinsic motivation. When an individual cannot manage the demands of the activity, or cannot exert control to obtain a desired outcome, amotivation will likely result (Ryan and Deci, 2020).

To achieve high-quality forms of motivation and engagement, three needs are seen as fundamental: autonomy, competence and relatedness (Ryan et al., 2019). Autonomy concerns a sense of initiative and ownership in one's actions. It is supported by experiences of interest and value, and undermined by experiences of being externally controlled, whether by rewards or punishments (Ryan and Deci, 2020). A large empirically-based literature has demonstrated the positive relations of more autonomous forms of classroom having more intrinsic motivation, perceived competence, and self-esteem (e.g., Deci, Schwartz, Sheinman, and Ryan, 1981), better grades (Vallerand, Fortier, and Guay, 1997) greater internalization for learning activities, and lower dropout (e.g., Hardre and Reeve, 2003; Vallerand, Fortier, and Guay, 1997). When students experience a sense of choice, they feel more ownership of activities and greater autonomy, resulting in an enhanced intrinsic motivation (Ryan and Deci, 2020).

Competence concerns the feeling of mastery, a sense that one can succeed and grow. The need for competence is best satisfied within well-structured environments that afford optimal challenges, positive feedback, and opportunities for growth (Ryan and Deci, 2020). Feedback can have informational significance if it is efficacy relevant (i.e., provides inputs that help the person improve or highlight areas of competence). Informational inputs tend to enhance intrinsic motivation and internalization. In contrast, feedback can have a controlling significance when experienced as pressure toward specific behaviours or outcomes (Deci and Ryan, 1985). Finally, relatedness concerns a sense of belonging and connection (Ryan and Deci, 2020). The thwarting of any of these three basic needs, possibly as a result of flawed learning activity design, is seen as detrimental to motivation. Accordingly, SDT's analysis of educational settings is primarily focused on the extent to which they meet or frustrate these basic needs (Proulx, Romero, and Arnab, 2017). Hence, we have focused on how an EEA embedding AR technology, as a tool for learning, can be developed around these fundamentals, to bolster engagement and learner outcomes.

Conditions supporting the individual's experience of autonomy, competence, and relatedness are argued to foster the most volitional and high-quality forms of motivation and engagement for activities, including enhanced performance, persistence, and creativity. In addition, SDT proposes that the degree to which any of these three psychological needs is unsupported or thwarted within a social context will have a negative impact on wellness in that setting.

## 3.5 ITMC Instrument Development

The determination of quantitative learning gains is notoriously difficult. The definition of learning gain adopted by McGrath et al. (2015) simply states it as the "distance travelled" by a student between two points in their academic career. As the core of this research is examining the impact of my AR technology-supported educational interventions on the student learning experience, an instrument was required as an attempt to determine learning gain as a consequence of the EEA. It is important to stress that I am not trying to author a Concept Inventory (CI). However, I do believe that the project benefited from the creation of an instrument containing items that could be used to quantitatively assess the following three principles of stereochemistry:

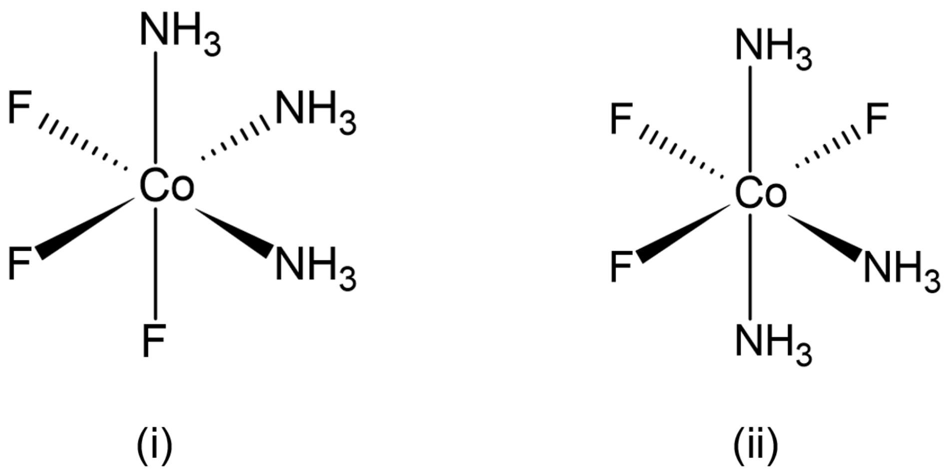
i. The rules of nomenclature.
ii. Stereoisomerism.
iii. Structural isomerism.

Although other stereochemistry CIs exist within the literature (Leontyev, 2015), existing instruments did not cover the scope of the principles that my intervention was developed around. Hence, an instrument was constructed for the purposes of this research, coined the Isomerism in Transition Metal Complexes (ITMC) assessment. The instrument contains 10 items in a multiple-choice format (figure 3.10), which are organised under the three outlined concepts important for developing proficiency: rules of nomenclature (items 1−3), stereoisomerism (items 4−6) and structural isomerism (items 7−10). Responses are scored as correct or incorrect (dichotomous), which are then aggregated to yield the total score.

A two-step validation approach was employed to ensure that the items on the instrument were appropriate to gauge conceptual understanding. Validation is the process by which a panel of experts are consulted to determine whether the instrument assesses what I intend it to assess. After creation of the initial draft of items, internal validation with experts in the field of inorganic chemistry at UEA was carried out. I asked each consulted expert to carefully read each item, and to see whether they agreed unambiguously with the selected answer, and to comment upon whether they agreed that the item was fit for purpose. This was carried out three times, with amendments being made to items where mutually agreed. Next, one round of external validation was carried out with experts from other UK universities. Changes were mostly attributed to the rewording of the stem of an item, or diagrammatic alterations. The output instrument, after the four rounds of internal/external validation, can be found in Appendix C.

.

6. Two isomers of coordination compound [$Co(NH_3)_3F_3$] are shown below. The isomers can be classified as:




a. (i) fac-isomer (ii) mer-isomer
b. (i) optical isomer (ii) trans-isomer
c. (i) mer-isomer (ii) fac-isomer
d. (i) trans-isomer (ii) cis-isomer

Figure 3.10. Item 6 on the developed ITMC test instrument.

## 3.6 AR Stereochemistry Escape Activity (2021–2022)

Educational design research is pragmatic, as it is concerned with the generation of usable knowledge, and solutions to challenges in practice. It uses theory to ground design choices, supported by empirical findings which guide changes made in a particular educational context, in accordance with emerging insights. This iterative process evolves through multiple cycles of design, evaluation, and revision (figure 3.11). Following the pilot study, principles of SDT were employed to drive design features of the second iteration of my stereochemistry EEA. My hope was that the narrative environment would serve as a context for students to enhance their understanding of stereochemistry concepts in coordination, whilst also developing their visual literacy (Hurley, 2022), and fostering intrinsic motivation. In contrast to the pilot study, this iteration of the activity was hosted digitally, and only incorporated elements of AR technology (neither molecular modelling kits, nor iVR technology were used). Like the pilot study, this stereochemistry EEA was integrated into UEA undergraduate module “Bonding, Structure and Periodicity”. Two further cross-sectional studies were carried out on this educational intervention, with two different student cohorts throughout the academic years of 2020/2021 and 2021/2022. One further iteration of development (which will be denoted as the third iteration) was conducted between these two cross-sectional studies.

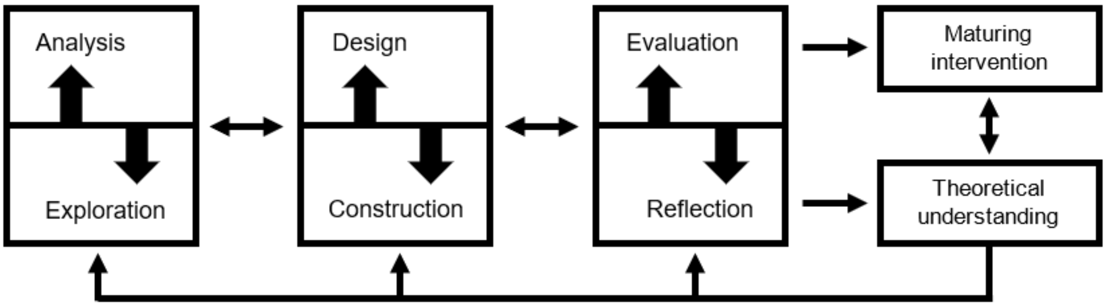



Figure 3.11. A model for conducting educational design research. Adapted from McKenney and Reeves (2012).

## 3.6.1 Experimental Design and Research Questions

This research design was carried out as a result of the identification of two main motives:

i. Relating to the motive of improving practice.
ii. Relating to the motive of enhancing the quality of research findings.

Throughout the cross-sectional studies, existing knowledge will be put to innovative use through the blending of AR affordances, and the pedagogical paradigm of the EEA. The experimental design employed is shown in figure 3.12. Like the pilot study, a pre-test/post-test design was employed, with participants randomly assigned to one of two groups to avoid bias and confounding variables:

i. **Experimental group 1** completed the EEA containing 2D structures of transition metal complexes. This group was treated as the control throughout the study.
ii. **Experimental group 2** completed the EEA containing embedded image markers for generating 3D virtual transition metal complexes.

Participants were assigned to one condition, either the control or AR-supported EEA, to eliminate carryover effects. The activity was structured as a 90-minute remote synchronous session composed of three sections: an introductory briefing, the EEA, and a debriefing session. Reflecting on the findings of the pilot study, and the reported limitations of the literature, several design changes were implemented.

Firstly, facilitating a physical EEA with large cohorts of students is difficult to achieve (Cain, 2019; Clarke et al., 2017). The creation of a digital EEA is not dictated by this constraint and allows hosting of large concurrent player bases. This is an approach better suited to large student cohorts typical of a university setting. The EEA was developed as a web-browser experience, as they are technologically undemanding, easy to modify, and are very accessible. The ability to utilise iVR technologies, such as the HTC Vive, was also inhibited due to the health and safety concerns with sharing hardware. This highlighted a distinct advantage of integrating AR technology for supporting immersive experiences. Students participating in the activity could easily download ChemFord onto their personal devices, allowing for feasible scalability, as well as the generation of virtual experiences from any location.

In preparation for this activity, a synchronous teaching session was conducted with the student cohort. This was composed of a 60-minute lecture on coordination chemistry. Students completed the ITMC test instrument at the pre-test phase to benchmark their understanding of the subject content, and at the post-test stage to allow for the completion of learning gain calculations. In addition to the ITMC, the Intrinsic Motivation Inventory (IMI) was also employed. The IMI is a multidimensional measurement instrument intended to assess participants' experiences in relation to a target activity. The instrument yields 7 sub-scale scores: (i) interest/enjoyment, (ii) perceived competence, (iii) effort, (iv) value/usefulness, (v) felt pressure and tension, (vi) perceived choice, and (vii) relatedness. The interest/enjoyment scale is considered to be the self-report measure of intrinsic motivation. Although the overall questionnaire is called the IMI, it is the only subscale that assesses intrinsic motivation (selfdeterminationtheory.org, 2022). The perceived choice and perceived competence concepts are theorised to be positive predictors of both self-report and behavioural measures of intrinsic motivation. Past research suggests that order effects of item presentation appear to be negligible, and the inclusion, or exclusion, of specific subscales appears to have no impact on the others (selfdeterminationtheory.org, 2022). As such, it is rare that all items on the IMI have been used in a particular experiment. Instead, experimenters have chosen the subscales that are relevant to the issues they are exploring. For this study, sub-scales (i), (ii), (vi), and (vii) were utilised. Previous application, and resulting analysis, of the IMI has shown strong support for its validity (McAuley, Duncan, and Tammen, 1989; Tsigilis and Theodosiou, 2003).

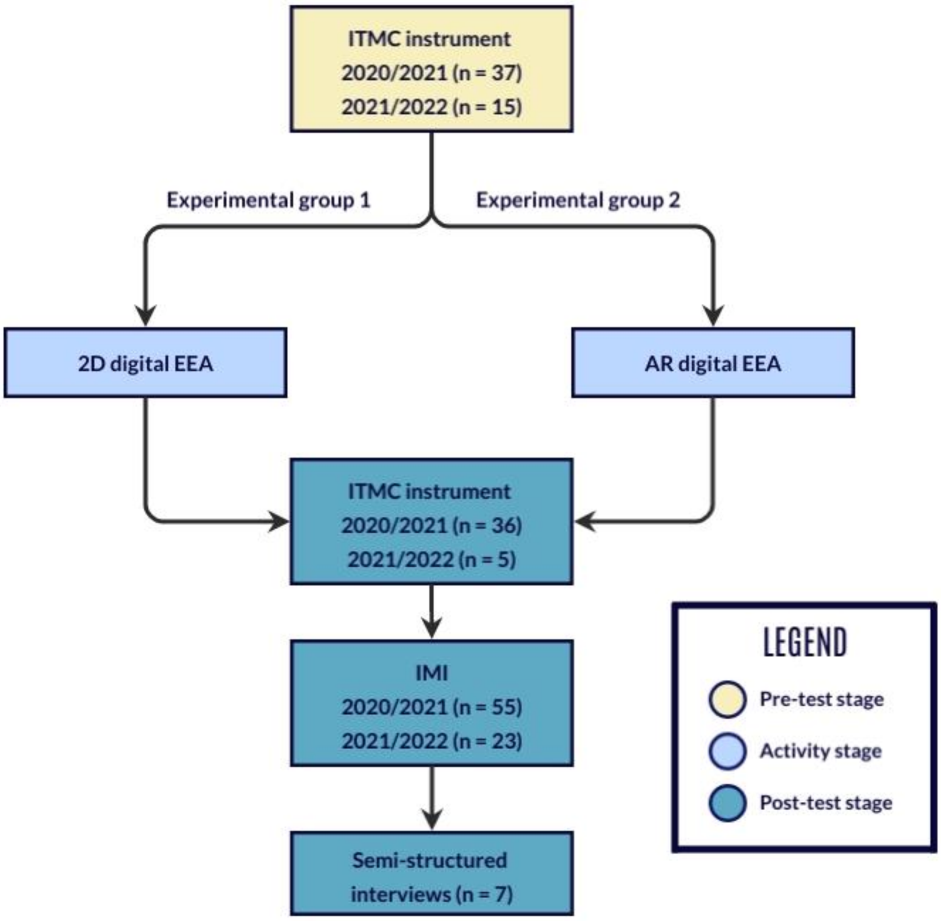



Figure 3.12. The experimental design utilised for this study, including details of participant engagement.

The cross-sectional studies presented in this chapter attempt to explore how an AR-supported digital EEA designed to support the psychological needs of autonomy, competency, and relatedness affects students' motivation, and their understanding of stereochemistry concepts. The research questions investigated throughout this chapter are as follows:

**Research Question 1a.** Do students who participate in the AR-supported EEA outperform students in the control condition on the ITMC test instrument?

**EEA Research Question 2a.** Are there significant differences between the two experimental groups regarding reported intrinsic motivation?

**EEA Research Question 3a.** What are the students' perceptions of the EEA as a learning experience?

### 3.6.2 Educational Escape Activity Design

Van den Akker (2006) suggests that the knowledge encompassed in design principles can be conveyed through the following heuristic statement:

If you want to design intervention *X*, for the purpose of *Y*, in context *Z*, then you are best advised to give that intervention the characteristics of *C1, C2... Cn [substantive emphasis]*. This is achieved via procedures *P1, P2... Pn [procedural emphasis].* This is because of theoretical arguments *T1, T2... Tn,* and empirical arguments *E1, E2... En*.

As such, the design aspects for each iteration of my stereochemistry EEA were informed by the framework of SDT, in relation to psychological needs satisfaction. In other words, how GBL elements could be implemented to ensure sufficient support of competency, autonomy, and relatedness. Students of autonomy-supportive teachers demonstrate greater learning outcomes (Vallerand, Frontier, and Guay, 1997), are more intrinsically motivated, and report higher perceived competence and internalisation of learning activities (Hardre and Reeve, 2003). I sought to actively support autonomy by providing a limited number of difficulty and exploration options. This was to avoid placing participants in a dilemma by offering too many choices. Further, I supported competency by integrating challenging, but achievable, tasks designed to the skill level of the students. On completion of a task, I integrated feedback mechanisms to positively inform players regarding their progress. Guidance on the stereochemistry principles being taught was delivered through the provision of support pages. These pages were accessible by players throughout the tasks to ensure that the challenges within the learning environment remained perceived as achievable. The focus was to clarify, and organise, content based on the knowledge and skills required to achieve the following learning objectives. Students should be able to:

i. Demonstrate application of the rules of nomenclature to create the name of a transition metal complex (in line with IUPAC recommendations, 2005).
ii. Differentiate different stereoisomers of transition metal complexes.
iii. Differentiate between different structural isomers of transition metal complexes.

Throughout the EEA, I attempted to facilitate social interaction, whilst eliminating factors that hinder the interactivity between students, to support their feelings of relatedness. When individuals feel they belong to a group, their need for relatedness is satisfied (Reeve, Deci, and Ryan, 2004). Due to the restrictions of the pandemic, peer-to-peer discussions were facilitated using Microsoft Teams (2022) breakout rooms, for each participating group of 3 players. Team-based tasks requiring contribution from multiple individuals were developed to foster collaboration. I also sought to evaluate the individual competency of each participant, in line with the learning objectives, within each participant group. This is an extension of previously utilised evaluation metrics in EEAs, such as completion rate, commonly used as an indication of competency among team members. To accomplish this, I introduced player roles, each with distinct sub-narratives and tasks that contribute to the team goal of completing the activity (figure 3.13).

It is noteworthy that the roles do not require any distinct prerequisite skillsets in relation to the other two roles. They were distinct in terms of narrative but covered the same underlying stereochemistry concepts. For example, the Agent role requires a student to apply principles of stereoisomerism to decrypt intel, whereas the Specialist role will apply the same principles of stereoisomerism to repair a reactor. Regardless of the role picked, each team member would encounter independent tasks, as well as collaborative team tasks, designed to promote proficiency in the topics of inorganic stereochemistry. The narrative used within the EEA was an extension of my previous pilot EEA, which aimed to collect qualitative data pertaining to students' experiences in this style of learning environment. The challenge thus evolved to constructing an AR-supported digital experience that incorporated the key competencies or learning objectives. The effectiveness of the EEA was examined using a mixed methods approach. Quantitative data regarding students' learning gains and measures of motivation were captured, alongside qualitative data pertaining to students' experiences.

To start the activity, once each participant had chosen their respective role within their team, the team were redirected towards a facility map, which acts as a hub for the tasks that require completion. The process flow for the first iteration of the digital EEA is shown in figure 3.14. Initially, most areas within the facility are inaccessible, but subsequent areas can be unlocked through completion of both individual and team-based (shared) tasks. The shared tasks are available immediately but require information from the role-specific tasks to complete. After the 2020/2021 student cohort had experienced the first iteration of my digital EEA, design changes were employed based upon discussion points identified during thematic analysis (section

3.6.6). The second iteration of my digital EEA contained the same tasks, but with a modified process flow, shown in figure 3.15. For example, Agent task 2, and specialist task 3 were changed from role-specific tasks to team-based tasks. Furthermore, Codebreaker task 3 replaced Agent task 3. This was done to improve the balance of shared tasks to role-specific tasks.

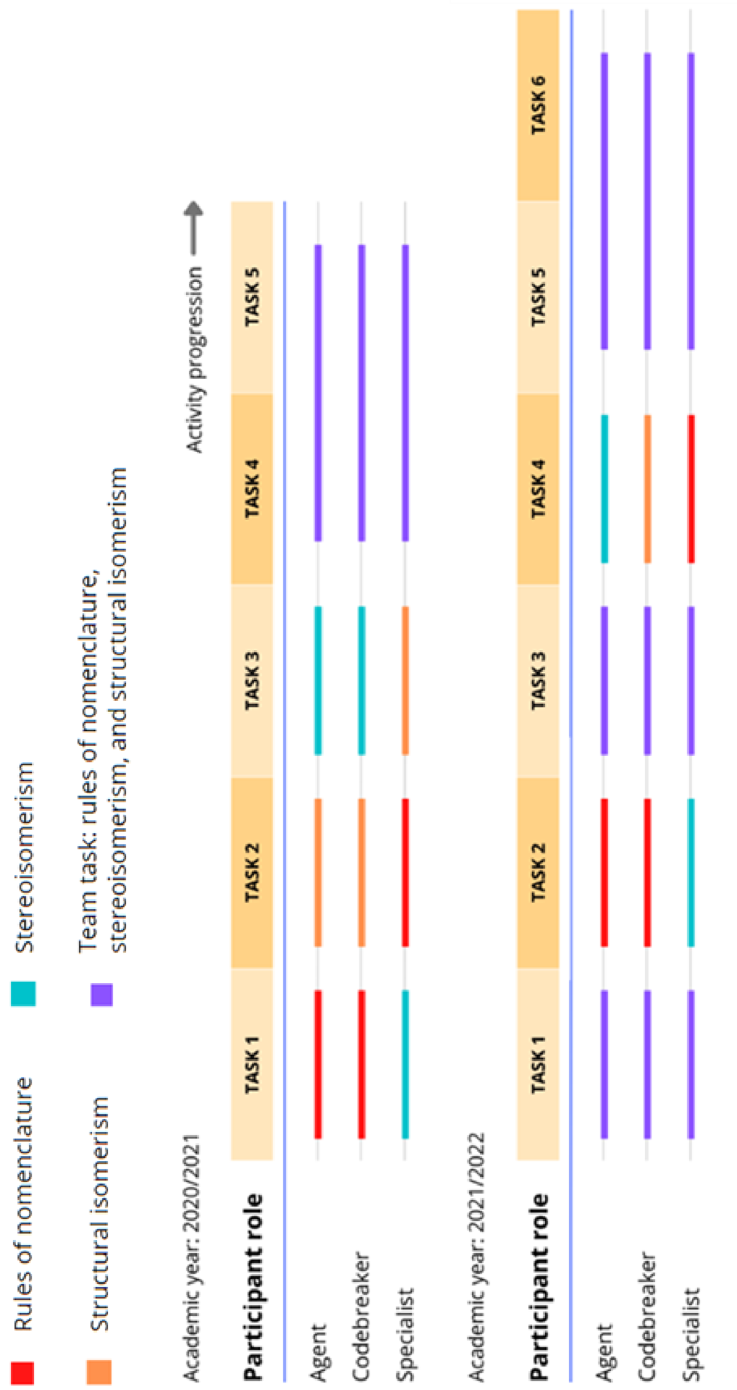



Figure 3.13. The sequence of tasks in the two iterations of my EEA, and their relation to each of the learning objectives.

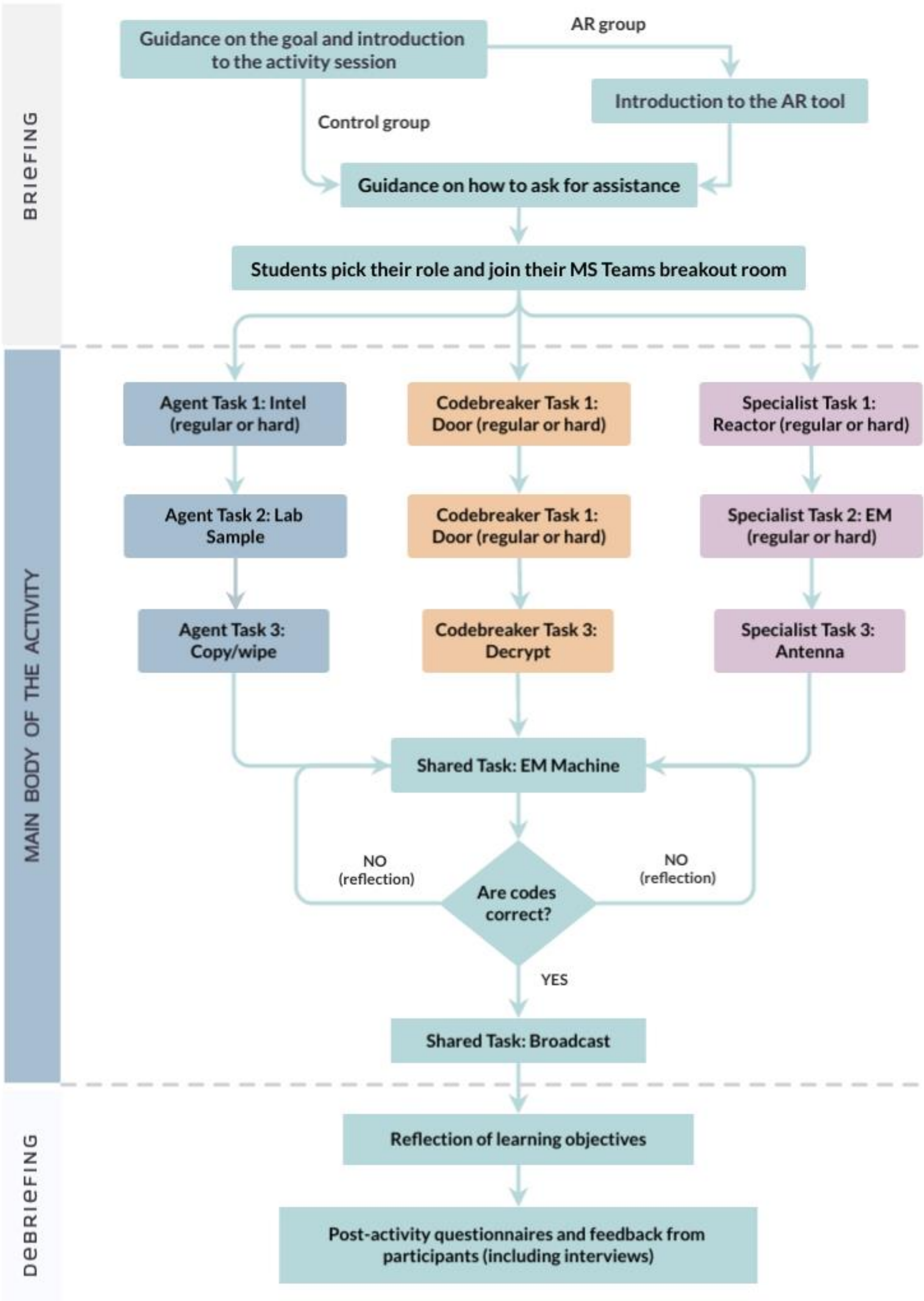



Figure 3.14. The process flow of the first iteration of the digital EEA, utilised during academic year 2020/2021.

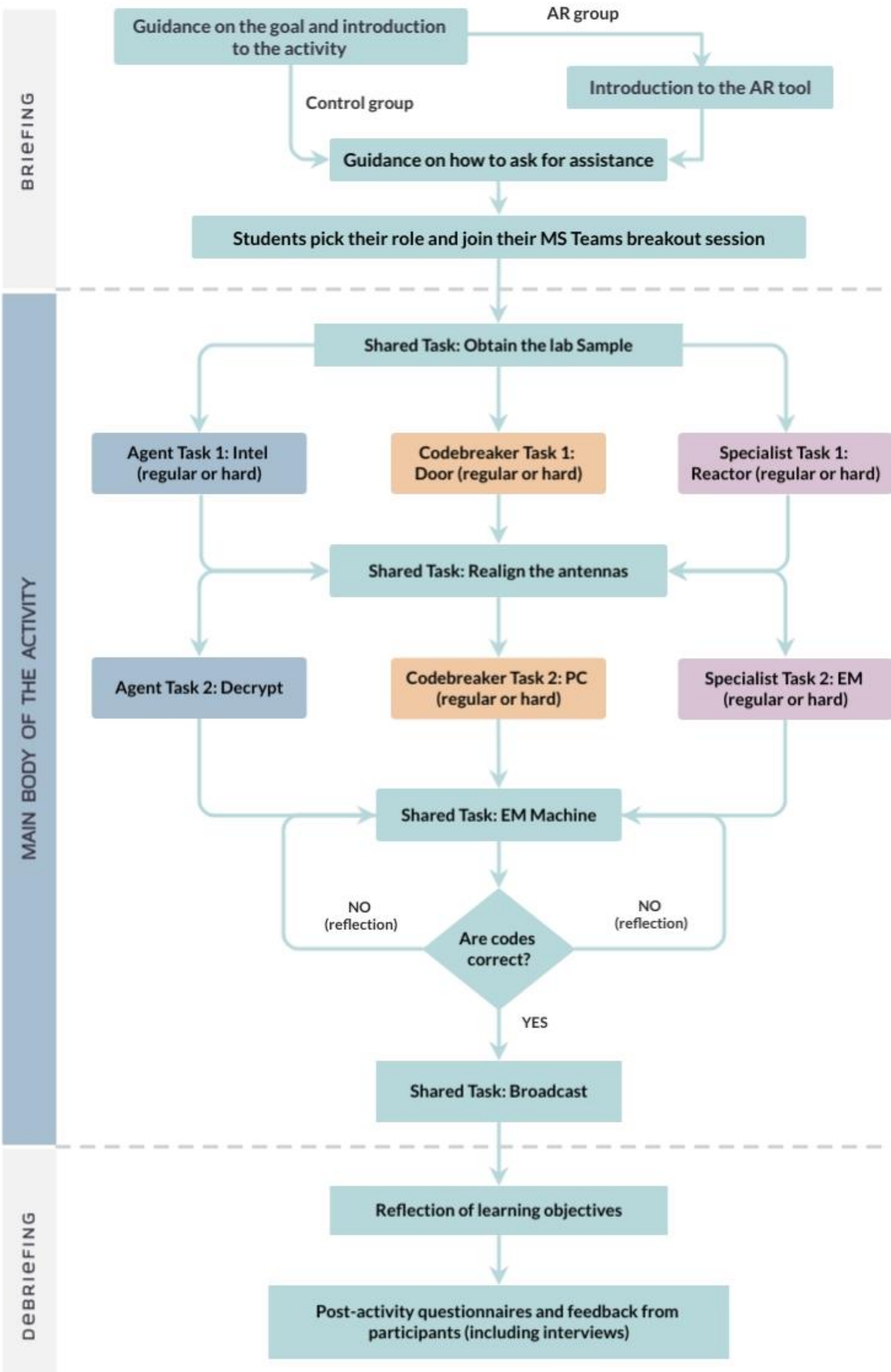



Figure 3.15. The process flow of the second iteration of the digital EEA, utilized during academic year 2021/2022.

### 3.6.3 Analysis of Participants' Performance

Table 3.6 outlines the descriptive statistics concerning the ITMC test scores achieved by participants prior, and in response to, completing my EEA intervention. Across the second and third iterations of the activity, 51 students completed the ITMC at the pre-test stage, and 40 students completed the ITMC at the post-test stage. Of these responses, 25 students completed the ITMC instrument at both the pre- and post-test stages. Following data collection, the Shapiro-Wilk test was used to check for the existence of normality. Although other methods for normality testing exist, Shapiro-Wilk has more power to detect the nonnormality on smaller samples sizes (Mishra et al., 2019). The data was found to be normally distributed at both pre- and post-test stages. In addition, Bartlett's test was conducted, verifying that the assumption of equal variances was true.

Table 3.6. Relative means and standard deviations for ITMC scores.

| **ITMC test instrument score**<br>***0 (low) to 100 (high)*** | **Control group**<br>Mean (SD) | **AR group**<br>Mean (SD) |
|---|---|---|
| Pre-test stage | 53.00 (15.67) | 43.33 (16.76) |
| Post-test stage | 79.00 (16.63) | 75.33 (20.31) |

Intergroup comparisons were conducted using the independent samples *t*-test. No significant differences were observed in the pre-test mean scores, $t(23) = 1.449$, $p = 0.161$. Therefore, it can be assumed that the two experimental groups were equal in terms of relevant chemistry experience. In addition, no significant differences were observed in the post-test mean scores, $t(23) = 0.474$, $p = 0.640$. Analysis of students' scores on the ITMC instrument demonstrated no significant differences between groups regarding performance on individual items. I hypothesised that the AR affordance of visualisation would result in participants from experimental group 2 performing better on items concerning stereoisomerism. This was not observed, $t(23) = 1.389$, $p = 0.178$. To measure intragroup performance on the ITMC, I utilised the paired samples *t*-test. Significant improvements were observed in ITMC test performance for both experimental group 1, $t(9) = 3.621$, $p < 0.01$, and experimental group 2, $t(14) = 4.262$, $p < 0.01$. Normalized change calculations were conducted as a measure of students' learning gain between the pre- and post-test stages. The $c$ values calculated were 0.44 for experimental group 2, and 0.50 for experimental group 1. To account for the variance in individual scores, I employed measures of effect size (Cohen's *d*) to compare differences between the groups in terms of learning gain. The suggested values for effect size were employed (Cohen, 2013): small (0.2), medium (0.5), and large (0.8). The calculated effect size was 0.14, meaning that the difference between the two groups was less than 0.2 standard deviations.

### 3.6.4 ITMC Instrument Reliability Analysis

A crucial component in the development of research instruments is establishing reliability, thus providing users with information regarding the quality of items. To better understand the item and scale characteristics of the ITMC instrument, I applied the concepts and analytical procedures of CTT and IRT. The association of CTT with basic statistical comparisons means that researchers who have had any exposure to measurement theory are likely to have encountered CTT. Thus, it can be utilised as a first step in establishing reliability. However, CTT has some noticeable shortcomings. Correlations being computed on the sample may differ between cohorts, and the methods employed do not involve the rigorous scrutiny of item characteristics that methods such as IRT employ. IRT models are non-linear monotonic functions describing the association between leaner ability on a latent variable and an item's characteristics on the probability of a particular response to that item (Embretson and Reise, 2000).

As in CTT, IRT requires that each item be distinct, yet consistent in reflecting the important aspects of the underlying construct. In the simplest case, IRT is evaluated in terms of one-parameter (1PL), difficulty, that determines the way an item behaves depending on learner ability. A two-parameter logistic model (2PL) introduces item discrimination, which determines the rate at which the probability of answering an item correctly changes with learner ability (Embretson and Reise, 2000). Lastly, three-parameter logistic models (3PL) introduce pseudo-guessing, which restricts the probability of endorsing the correct response as ability approaches $-\infty$.

Figure 3.16 shows the calculated properties of difficulty and discrimination for items on the ITMC using CTT. In the context of educational testing, a difficult item is one that more respondents answer incorrectly. The difficulty values calculated range from 0–1, where a higher value indicates an easier item. The most effective items have mid-ranges of difficulty. However, in practice, a difficulty of 0.5 on every test item for every cohort is not feasible. Therefore, difficulty values within a range of 0.3–0.9 are acceptable. Items more strongly correlated with other items, and thus the true score, are fundamentally better items. Such items are said to have greater discrimination. The extreme group method was used to calculate discrimination with groups partitioned by the top and bottom 27% (Preacher, 2015).

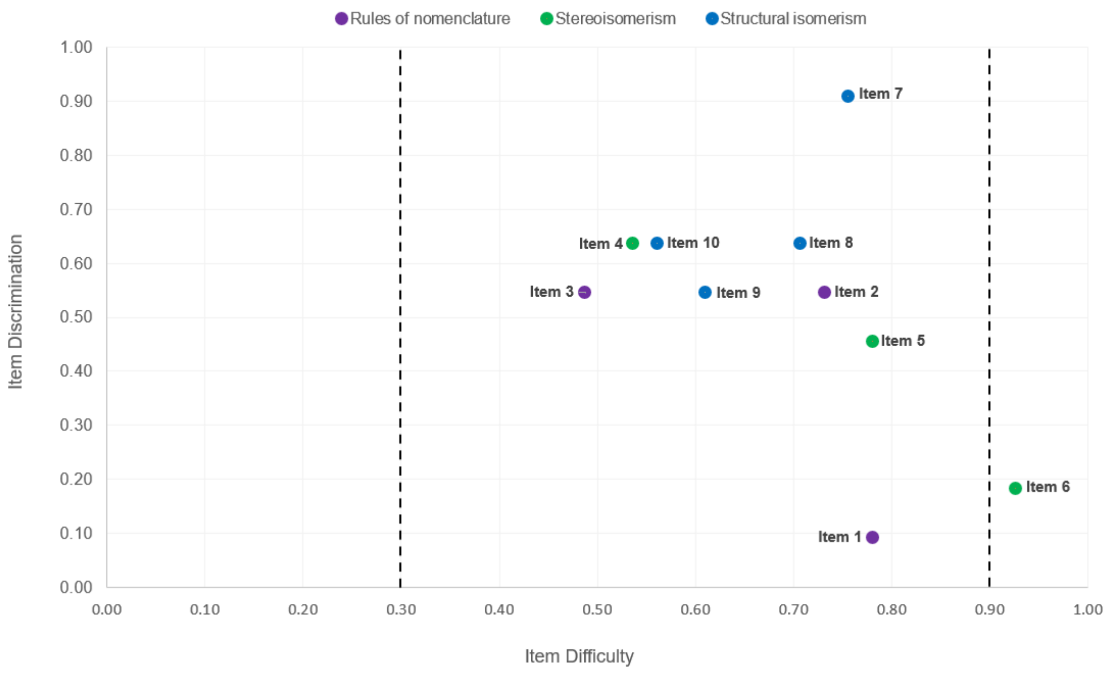



Figure 3.16. Item difficulty and discrimination values for each of the 10 items in the ITMC test instrument. The black lines represent the recommended upper and lower bounds of item difficulty. Each dot represents an item.

IRT analysis of the ITMC data was performed using dichotomous 1PL, 2PL, and 3PL models. To assess the absolute fit of each model, two measures were examined. Firstly, a generalisation of Orlando and Thissen's (2003) S-$\chi^2$ item-fit statistic was inspected. The item-fit statistic assesses the degree of similarity between model-predicted and empirical response frequencies by item response category. A statistically significant value indicates that the model does not fit a given item. The S-$\chi^2$ fit statistic for each item (table 3.7) indicates a satisfactory fit in 8 of the 10 items for the 1PL model. For the 2PL model, no items displayed a non-satisfactory fit. Addition of the pseudo-guessing parameter (3PL model) introduced a non-satisfactory fit in items 4, 7, and 9.

Table 3.7. Item-fit statistics for 1PL, 2PL and 3PL IRT models

| ITMC item number | 1PL | | 2PL | | 3PL | |
|---|---|---|---|---|---|---|
| | ***S-χ2*** | ***p*** | ***S-χ2*** | ***p*** | ***S-χ2*** | ***p*** |
| 1 | 12.94 | 0.044 | 4.32 | 0.504 | 3.56 | 0.468 |
| 2 | 2.71 | 0.844 | 2.68 | 0.749 | 4.41 | 0.353 |
| 3 | 5.12 | 0.529 | 4.26 | 0.513 | 6.10 | 0.192 |
| 4 | 10.51 | 0.105 | 9.31 | 0.097 | 11.58 | 0.021 |
| 5 | 5.58 | 0.472 | 5.43 | 0.366 | 7.82 | 0.099 |
| 6 | 6.60 | 0.359 | 5.77 | 0.329 | 8.63 | 0.071 |
| 7 | 14.14 | 0.028 | 9.74 | 0.083 | 12.39 | 0.015 |
| 8 | 4.47 | 0.614 | 4.30 | 0.508 | 7.47 | 0.113 |
| 9 | 11.26 | 0.081 | 10.08 | 0.073 | 13.52 | 0.009 |
| 10 | 3.54 | 0.739 | 3.69 | 0.594 | 4.10 | 0.392 |

The fit of the 2PL and 3PL models to the data were also compared using the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) fit statistics (table 3.8) (Acquah, 2010). For the two statistics, a lower value indicates a better model fit to the data. The addition of the pseudo-guessing parameter (3PL model) did not improve the model fit. Thus, the 2PL model was used to interpret item parameters.

Table 3.8. Model level fit comparison for 2PL and 3PL models for this study.

| **Model** | **Log-Likelihood** | **AIC** | **BIC** |
|---|---|---|---|
| 2PL | −187.51 | 407.02 | 434.44 |
| 3PL | −185.43 | 418.86 | 459.99 |

Figure 3.17 displays the item-characteristic curves (ICC) generated from my 2PL model. ICCs are the fundamental unit in IRT and can be understood as the probability of answering a dichotomous item correctly, for individuals with a given ability (Embretson and Reise, 2000). Items that are easy to correctly answer are shifted to the left of the scale, whereas items that are difficult to answer correctly are shifted to the right. Generally, the ICC have an ogive curve, beginning on the left with low probabilities of answering an item correctly for lower values of student

ability, rising to represent increasing probabilities of answering the item correctly as student abilities increases.

The item threshold (the point at which curve inflection occurs) indicates the item's difficulty. For the 2PL model, the inflection point occurs where the curve crosses the median probability value. This indicates the student ability for which the probability of answering the item correctly is 0.5. In addition, items also have an estimated discrimination parameter allowing item curves to have different slopes. The steeper the slope of the item response function, the better that item discriminates among students of different abilities. Students whose ability measures are at the flatter ends of the item's ogive curve cannot be separated with a great degree of confidence by that item. Items on the scale displayed good discrimination, constituting reasonable evidence that each item's score is positively related to the overall proficiency represented by performance on this instrument. Items 1 and 6 were considered the easiest items, generally at the lower estimate of individuals' ability. Item 6 was also found the be the easiest scale item when analysing the ITMC using CTT. This is represented by the ICC generated from my 2PL model. The inflection points of items 1 and 6 lie at an ability lower than −4. As such, I have omitted them from the item characteristic curves shown in figure 3.17.

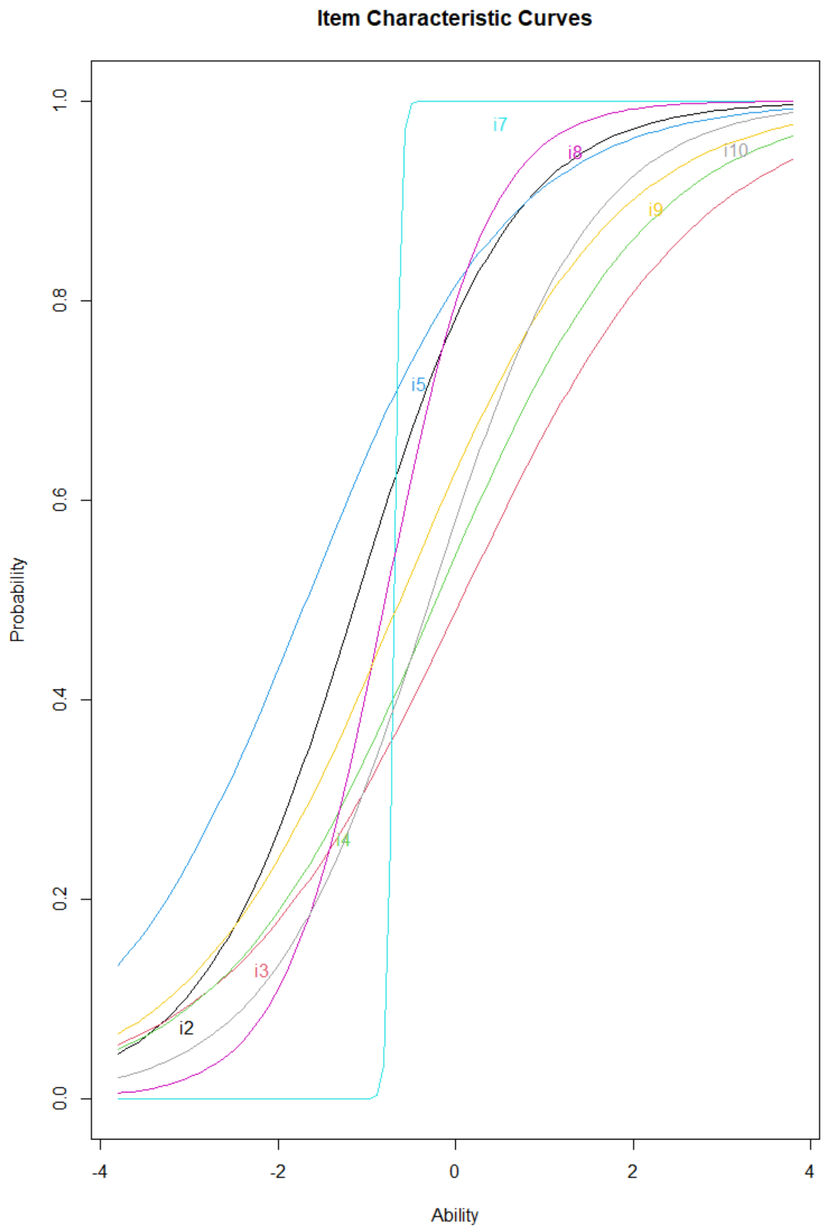



Figure 3.17. The item characteristic curves for items on the ITMC, generated using a 2PL model, excluding items 1 and 6.

I employed Differential Item Functioning (DIF) to see if students of equal ability, but from different experimental groups, had unequal probability to respond correctly to the items on the ITMC instrument (table 3.9). This is because DIF items can lead to biased measurement of ability. The DIF is stated to be uniform or non-uniform depending on whether the discrepancy in item performance between subgroups is consistent or non-consistent respectively. Raju Signed Area method was employed, using detection thresholds of −1.96 and 1.96, with a significance level of 0.05. No items on the ITMC test instrument were detected as DIF items (figure 3.18).

Table 3.9. Differential Item Functioning for items on the ITMC test instrument.

| **Item Number** | **Raju Signed Area method** | | |
|---|---|---|---|
| | Statistic | *p* value | DIF detected? |
| 1 | 0.137 | 0.891 | NODIF |
| 2 | -0.439 | 0.661 | NODIF |
| 3 | -0.130 | 0.896 | NODIF |
| 4 | -0.665 | 0.506 | NODIF |
| 5 | 0.298 | 0.766 | NODIF |
| 6 | -0.654 | 0.514 | NODIF |
| 7 | 0.018 | 0.986 | NODIF |
| 8 | 0.793 | 0.428 | NODIF |
| 9 | 0.584 | 0.559 | NODIF |
| 10 | 0.747 | 0.455 | NODIF |

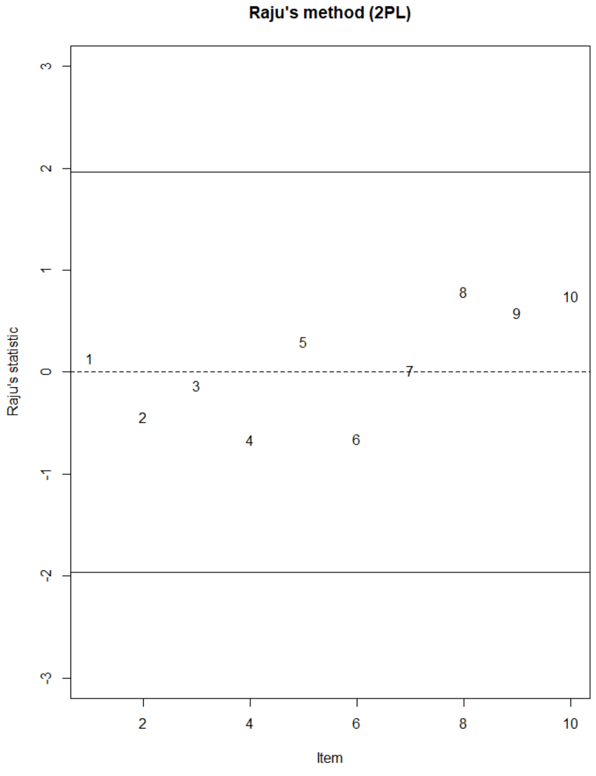



Figure 3.18. Output of Raju's method (2PL) performed on ITMC response data.

### 3.6.5 Participant's Measures of Motivation

The descriptive statistics pertaining to students' responses on the IMI are presented in table 3.10. The authors of the original scale encourage adaption of items for use with different populations and in specific activities (selfdeterminationtheory.org, 2022). As such, the internal consistency of the instruments' sub-scales was established through calculation of Cronbach's alpha. The computed values (shown in table 3.10, column 5) are indicative of good internal consistency. Item deletion procedures suggested a higher alpha-if-deleted value for one item on the relatedness sub-scale: *(item 25: I'd really prefer not to interact with this person in the future)*. No item on any of the other three sub-scales demonstrated a higher alpha-if-deleted value. Intergroup comparisons for each of the IMI sub-scales employed was conducted using the Mann-Whitney U test, a non-parametric test for ordinal data. The calculated asymptotic significances show that self-report measures of intrinsic motivation from participants in experimental group 2 were not significantly different to those reported by experimental group 1.

Table 3.10. Results from the IMI presented as median (interquartile range).

| IMI sub-scale (7-point Likert scale) | Control group (n = 38) | AR group (n = 40) | Asymp Sig. (2-tailed) | A |
|---|---|---|---|---|
| Interest/Enjoyment | 5.14 (1.39) | 5.00 (2.21) | 0.766 | 0.909 |
| Perceived Competence | 4.00 (1.87) | 3.75 (2.67) | 0.306 | 0.943 |
| Perceived Choice | 4.57(1.57) | 4.93(2.46) | 0.714 | 0.869 |
| Relatedness | 4.94 (1.44) | 5.00 (1.38) | 0.715 | 0.748 |

In addition, Spearman's correlations were conducted to explore the relationships between the constructs reported by each sub-scale of the IMI and the ITMC instrument (table 3.11). The interest/enjoyment sub-scale was strongly correlated with the perceived choice sub-scale, and moderately correlated with the perceived competence sub-scale at the $p = 0.01$ level. This agrees with the hypothesis that perceived choice and perceived competence are positive predictors of measures of intrinsic motivation, considered to be assessed by the interest/enjoyment sub-scale. The perceived choice sub-scale was moderately correlated with the perceived competence sub-scale at the $p = 0.05$ level. The relatedness sub-scale did not display significant correlation with any of the three other IMI sub-scales. No significant correlations were observed between ITMC test scores and the four endorsed IMI sub-scales.

Table 3.11. Spearman's correlations conducted between IMI sub-scales, and between IMI sub-scales and the ITMC test instrument.

| Measure | $r_s$ | | | |
|---|---|---|---|---|
| | **Interest/ Enjoyment** | **Perceived Competence** | **Perceived Choice** | **Relatedness** |
| Interest/Enjoyment | 1.000 | 0.698** | 0.489** | 0.028 |
| Perceived Competence | 0.698** | 1.000 | 0.388* | 0.091 |
| Perceived Choice | 0.489** | 0.38* | 1.000 | 0.189 |
| Relatedness | 0.028 | 0.091 | 0.189 | 1.000 |
| ITMC total score | 0.163 | 0.016 | 0.001 | −0.070 |

* Correlation is significant at the 0.05 level (2-tailed) ** Correlation is significant at the 0.01 level (2-tailed)

### 3.6.6 Analysis of Qualitative Data

I recruited 7 students in total, from both experimental groups, to participate in semi-structured interviews. The interview schedule (Appendix D) covered four topic areas:

i. Perception and satisfaction in response to attempting the EEA.
ii. Interest and experience with games.
iii. Value and usefulness of the EEA.
iv. Activity pressure and effort.

Qualitative analysis of the participant interviews was completed through latent thematic analysis using the approach of Braun and Clarke (2006). Data was recorded, and transcribed verbatim, prior to being subjected to analysis for commonly occurring themes. The initial broad themes were constructed based on frequency and similarity of responses. Redundancy was eliminated and closely related major themes were merged. For this research, I focused on three predominant themes found in student discussions: application of the subject content, affective and motivational factors, and evolving the activity. Study interviewees (SI), for purposes of pseudonymisation, are again represented by a number.

I sought to ensure reliability in my analysis using negotiated agreement. The extent of agreement between coders was measured using Krippendorff's alpha. Two researchers independently coded the full set of interview transcripts and then negotiated in how they applied the codes. Differences were discussed and where there was a consistent disagreement, a common approach was agreed (the negotiated codebook employed can be found under Appendix E). Krippendorff's alpha is a commonly used chance-corrected reliability measure that avoids many of the limitations described for Cohen's kappa, such as its suitability to smaller samples sizes (Krippendorff, 2018). Krippendorff's alpha has ranges between −1.00 and 1.00, with positive values indicating agreement beyond chance. Values above 0.66 are acceptable for tentative conclusions (Krippendorff, 2018). The Krippendorff's alpha calculated for this set of coded interview transcripts was 0.84.

The first theme pertains to the **application of the subject content**. All participants expressed views on the difficulty of the EEA in terms of both the game mechanics and the embedded chemistry content. Supporting the need for competency, students could attempt the same tasks at different levels of difficulty. Many students stated that the difficulty of the activity was suited to their level of chemistry experience, supporting the need for competency. *"I don't think it was easy, but it wasn't too hard either. I think it was the right amount of challenging."* (SI 7). Of all participants who attempted the activity, 42% of those selecting the specialist role attempted hard difficulty challenges. Further, 45% of codebreakers and 39% of agents also attempted the hard difficulty challenges. Participants articulated that to improve, *"…it needs to be challenging, at least to a certain extent, for it to change you in a better way…"*, (SI 6).

Within this theme, I can identify different aspects relevant to learning. Participants stated that design aspects such as support pages "reinforced" the learning content throughout the activity. Responses suggest students found the activity to be a meaningful learning experience, "*I got something out of it. When I did the test, I got 8 out of 10, and I don't think I would have if I hadn't done the activity. It reinforces a lot of things."* (SI 2). This supports my collected quantitative data. Paired sample *t*-tests, and normalized change calculations, demonstrated significant intragroup improvement on the ITMC test instrument prior to, and after, the activity (table 1 and figure 7). Within my discussions, students demonstrated reflection, *"I realized where*

*I needed to go back and look…"*, (SI 5), and stated that the opportunity to apply taught content promoted deeper understanding of the material.

*"It got me to read those notes again, to facilitate answering these questions, and I thought that was really reinforcing."* (SI 7).

*"I sit in lectures thinking I understand the context of the chemistry at the time but having to use it in a different way immediately afterwards helped reinforce it."* (SI 1). My qualitative data indicates that the introduction of ChemFord into my EEA did not result in significantly higher post-test results on the ITMC test instrument, compared to the control EEA condition. I hypothesised that ChemFord would assist cognitive processing associated with mental visualisation, thus supporting students, for example, when approaching problems regarding the spatial relations of ligands in transition metal complexes. However, variables such as intrinsic and extraneous cognitive load were not measured as part of this study. Conducting an ANCOVA shows no significant differences between the different aspects of the ITMC test instrument between groups: rules of nomenclature, $F(1,24) = 0.516$, $p = 0.480$, stereoisomerism, $F(1,24) = 0.452$, $p = 0.508$, and structural isomerism, $F(1,24) = 0.071$, $p = 0.792$.

The second theme of this thematic analysis is in relation to **affective and motivational factors**. In their accounts, participants highlighted their experiences of the EEA. During a challenging period of transition to online learning, students positively perceived the integration of my online synchronous activity. *"I really enjoyed it. I thought the escape room was really well made. I thought it was really good fun."* (SI 1). This was supported by higher reported measures of interest/engagement on the IMI survey. When asked, students expressed a desire to repeat this style of activity in future modules throughout their degree. *"I would definitely want to see it happen again, not just in this module, or in this course, I'm sure it's going to be beneficial for other courses as well."* (SI 6). Participants frequently used terms such as "engaging", "satisfying", and "useful" to describe the activity.

*"I'd say it was a good use of time to consolidate things and correct some misconceptions that I had beforehand."* (SI 7).

In contrast, negative student feelings were also noted. An absence of direct instruction, due to the nature of the activity, left some students feeling initially overwhelmed. *"There was definitely some stress and anxiety at the start"* (SI 4). In addition, students expressed that they felt the time pressure. However, most students stated that *"…when we started bouncing ideas off each other on how to progress, that anxiety started to go away…"* (SI 7), and that it became *"…more enjoyable than stressful…"* (SI 3). On supporting the need of relatedness, students positively responded to collaboratively working given the limited interaction with their peers.

*"It was nice to have to work with someone from the course. Because this year, I haven't really met many people from the course."* (SI 6).

All participants stated that they believed the EEA worked as an effective team activity. Students also expressed support towards their peers, *"…when a person in the team competed their part… I don't know if proud is the right word?"* (SI 3). However, challenges regarding the facilitation of the team interactivity were raised.

*"If possible, [do the activity] in person next year. And that's always better, because it's so much easier to get past that initial awkwardness in person than it is online."* (SI 6).

*"I would say, I think if I had been in a group where I didn't know anybody, I probably would have felt anxious about meeting them, and having to speak."* (SI 1).

Evidence of extrinsic motivation was apparent, *"...the other two members had taken the effort to show up. They needed codes from me to complete it…"* (SI 1), with another participant exclaiming that *"…trying to find that intrinsic motivation is quite challenging for me…"* (SI 3). To explore the topic of motivation, we asked participants about their gaming experience outside of an educational setting. Most students stated that they "play a lot of video games", but with no preference for competitive or cooperative play. Following this, my discussions led onto what motivated participants to continue playing a game once the difficulty surpasses their current ability. Responses typically fell into (i) competitiveness, *"I think it appeals to my competitiveness",* (SI 1), and (ii) self-improvement, *"Improving myself in order to feel like I'm good enough."* (SI 5). To understand if this translated to the EEA, I posed a similar discussion with my participants. Self-improvement and the contribution of the activity to participants' learning were the primary responses. *"For me, cracking the safe and completing the puzzle is a reward in itself. I want to know that I can do it."* (SI 2).

Between-groups, the introduction of ChemFord did not result in greater measures on any of the four sub-scales of the IMI. Qualitative discussions show evidence of extrinsic motivation, more specifically the process of identification. This is represented by students showing conscious valuing of the activity, and personal importance. The process of internalisation towards intrinsic motivation is also evident, with students stating their interest, enjoyment and inherent satisfaction of the learning experience. However, it is noteworthy that these perceptions were present in both experimental groups, not just those students utilising the AR technology. I hypothesised that introducing ChemFord as an educational tool would help to support the psychological need of competency. Yet, measures of perceived competence between the two groups were not significantly different, $t(76) = 1.070$, $p = 0.288$, but were shown to be positively correlated with intrinsic motivation. Again, the difficulty of the tasks within the EEA were perceived as sufficiently difficulty by both groups. As such, the introduction of AR may not have provided the cognitive benefits I perceived it would in this instance. Furthermore, there was a risk that AR may have thwarted the need for autonomy through requiring the user to interact with ChemFord specifically. However, perceived choice was not significantly different between groups, $t(76) = 0.267$, $p = 0.790$. Lastly, qualitative data did not provide any evidence of amotivation or external regulation.

The last theme emerging from my thematic analysis is **evolving the activity**. As a formative session, engagement in the EEA was a choice by my participants. Therefore, an important consideration is to identify how to provide supportive strategies to ensure students will be more likely to experience psychological need satisfaction. SDT lends itself well to intervention work, as throughout the iterative design process, we can capitalize on the opportunity to explore whether improving the levels of need support in my EEA positively impacts levels of intrinsic motivation and the targeted learning outcomes.

Several discussion points were captured for consideration between the second and third iterations of the EEA. Firstly, participant teams will commonly be composed of students of differing relevant chemistry experience. As such, instances arose with individual tasks, where players were completing them at different rates. This resulted in the generation of “dead zones” where players were potentially inactive while awaiting further information from their teammates.

*“I finished my tasks before the other two did. So, all I was doing was it was helping my teammates do their tasks. If I’m being honest, I sat there thinking, ‘okay, I need something to do whilst I’m waiting’”*. (SI 1).

This is an example of a relatedness thwarting strategy, exhibited by an active dislike towards an aspect of the learning environment. To support relatedness and inclusion of team members, the second iteration of my EEA was designed to begin with a team task, as well as having a lower emphasis on individual tasks. This design idea was suggested by students throughout the qualitative data collection, *"100%. I think using a team task at the start to help people bond straight away would be a good idea."* (SI 4). Whilst facilitating the EEA activity in academic year 2021/2022, it was apparent that greater levels of peer-to-peer discussion were taking place as a result of this design change. A greater measure of relatedness on the IMI survey was reported during the second iteration (5.44), but this was not statistically significant when compared to the first iteration.

Balancing the GBL mechanics with the chemistry content of a task was also commented upon, "*...there were a couple of points where, I think, individually, we were a bit stuck to the premise of a couple of the tasks, and exactly what it wanted from us, rather than the chemistry*", (SI 1). To avoid thwarting the need for competency, we reviewed data from two sources to inform whether tasks required revision between iterations of the activity: (i) qualitative feedback from participant interviews and, (ii) quantitative data gathered from tracking statistics. Where participants explicitly stated that a task was difficult, or tracking statistics displayed minimal player progress, changes were made to ensure tasks remained achievable. This also avoids the GBL elements confounding with the potential benefits of the AR technology.

## 3.7 Limitations

The research studies presented in this chapter, evaluating both the pilot EEA and subsequent iterations of the digital EEA have limitations that must be discussed. Firstly, a major limitation is the relatively small sample size that the data analysis was based upon. Due to the window to operate the activity, the number of participants with the opportunity to engage in this educational intervention was limited. The sample size was the result of modest enrolment, compounded by participant disengagement between the pre- and post-test stages. As such, it is not possible to generalise my findings based on the sample size of the two studies. Secondly, following the adoption of online learning in response to the COVID-19 pandemic, there was no opportunity to observe students' interactions with the AR technology when participants were completing the EEA. It would be interesting to understand how IMI measurements for students interacting with ChemFord alone compare to those of students engaging with my AR-supported EEA. Thereof, only a small amount of observational data was collected, predominantly from the pilot study.

Furthermore, the feasibility assessment of the EEA was carried out by a single individual, an exercise that must be conducted with a larger group of facilitators. To further evaluate the learning potential of the EEA, repeat activities are required with larger cohorts of participants. Furthermore, we must acknowledge the possibility of self-selection bias from participants (Heckman, 1990). Students who volunteer for interviews may be different from the rest of the population regarding their communication ability or reasoning levels. We were unable to evaluate the learning gains of students who did not participate in either the control or AR condition (i.e., no EEA intervention). This would allow me to understand if a student who completed the ITMC test instrument twice displayed significant improvements in their score, as a result of reflection between the pre- and post-test stages. Lastly, the 2D nature of the ITMC test instrument may introduce bias towards the control condition. It is unknown at this point whether an AR version of the ITMC would influence test performance.

## 3.8 Chapter Conclusions

There is little doubt that motivation can be influential on the degree to which individuals engage with learning experiences. This chapter reports the implementation of an EEA for supporting higher education chemistry students' understanding of stereochemistry principles. A pilot studied was initially conducted, as a novel approach to comparing EEAs implementing different immersive technologies. With the implementation of AR and iVR software into the EEA, it was critical to differentiate between software errors and task elements, as this can cause confusion within the whole experience. The potentially significant technical expertise required to develop, include, and maintain interactive computer-based systems is currently considered a major drawback. For successful implementation, facilitators will require the skills to recognize and troubleshoot problems, whilst developers will need to carefully test applications, to ensure high educational value. The initial results of the pilot study indicated the effectiveness of using an EEA, in terms of student engagement, and found that students valued educational tools such as AR and iVR when learning stereochemistry concepts. One of the main questions raised by the pilot study is how the EEA can be utilised to evaluate participants for competencies and skills pertaining to the learning objectives. Following the pilot study, the second iteration of the intervention retained the core foundation provided by the escapED framework, but also integrated design features based on the theoretical underpinnings of SDT. The design approach of a digital EEA provided many positive results, implementing technical elements which can be easily upscaled for larger cohorts of concurrent players. According to the data provided in this work, EEAs promise to be a valuable contribution to higher education chemistry teaching, and merit further research in the educational community.

The second and third iterations of this intervention illustrate how the design of an AR-supported EEA, to support motivation, can be employed. The design of this EEA provides one approach to implementing this style of GBL activity, in a way that supports virtual presence, and is scalable to large student cohorts. Collected qualitative data suggests that participants found the activity to be useful and engaging. Examples of extrinsic motivational factors were mentioned by interview respondents. Further work examining the effects of intrinsic and extrinsic motivation on students' perception and performance in educational escape initiatives would be a welcome addition. Through students' discussions, I have provided evidence of how design aspects of the EEA support the psychological need satisfaction outlined by SDT. This indicates how future evolution of the activity can address these needs. With reference to the research questions, the introduction of AR, over and above the EEA, did not result in any significant differences in reported intrinsic motivation, or post-test scores on my stereochemistry instrument. Initial reliability evidence for the ITMC test instrument, which was developed for the purposes of this study, has been provided using the approaches of CTT and IRT (2PL model). Items 1 and 6 were shown to be the easiest items on the scale. DIF showed no biased measurement of ability between groups when using the Raju Signed Area method. Reported measures of competency were seen as a positive predictor of intrinsic motivation. However, in this study, this was not observed to be a positive predictor of academic performance. Significant intragroup academic improvement was observed in both experimental groups.

Despite the growing interest in educational games for learning, further empirical evidence is necessary to evaluate the potential of GBL with regards to both the learner and game design. According to student feedback and observations, the EEA was an engaging experience. Finally, future work should concentrate on defining the game mechanics most appropriate to addressing both the pedagogical and learning objectives. Given that there is limited work on serious gamification through the utilisation of escape rooms in chemistry higher education, and even less incorporating immersive technologies, there is a potential to gain considerable information about the development of key chemistry competencies using this teaching strategy. Research regarding how key elements of the EEA, such as the briefing and debriefing sessions, should be designed and constructed is yet to yield definitive insights. The debriefing session is important as a time of reflection on the learning objectives and to provide feedback.

# 4

# AR-Supported Game-Based Learning for teaching VSEPR

The Valence Shell Electron Pair Repulsion (VSEPR) theory is an archetypical example of stereochemistry. It is a model in chemistry that provides an explanation for the basic geometry of many main group compounds encountered by higher education chemistry students based upon the extent of electrostatic repulsion. The "AXE" method of electron counting is commonly applied to determine the shape of a molecule based on the principles of VSEPR (Burrows et al., 2021):

- The "A" represents the central atom.
- The "X" represents *m* number of bonds between the central atom and its substituents.
- The "E" represents *n* number of lone pairs surrounding the central atom.

The sum of *X* and *E,* obtained from a molecule's Lewis structure, are denoted as the steric number. In $AX_mE_n$ molecules, electrostatic interactions repelling volumes of negative charge leads to the formation of a most-probable octahedral shape to maximise the distance between the fluorine substituents to reach an energetic minimum (Gillespie, 1963). Visualising three-dimensional (3D) shapes requires cognitive processes in the spatial domain, and thus, it is crucial that students can mentally perceive them. Consequently, educators are increasingly introducing a variety of instructional media and resources to teach the principles of VSEPR. Previous works reported in the literature include approaches designed around Game-Based Learning (GBL) and molecular model building (Erlina, Cane, and Williams, 2018), molecular computer modelling and use of experimental data (Martin, Vandehoef, and Cook, 2015; Pfennig and Frock, 1999), and 3D printing technology (Dean, Ewan, and McIndoe, 2016).

In chapter 4, the development and evaluation of my second augmented reality (AR)-supported educational intervention is discussed. I have called this intervention "The

City of Gillespie”. As discussed in chapter 3, the paradigm of GBL presents many advantages for use in educational settings in terms of stimulating motivation, increasing interest, and promoting active involvement with the learning activity. As such, like intervention 1, my second intervention was developed with GBL as the pedagogical underpinning. Yet, where intervention 1 used the framework of Self-Determination Theory (SDT) to guide design features, intervention 2 were underpinned by theories such as Cognitive Load Theory (CLT) and the Cognitive Theory of Multimedia Learning (CTML). As such, variables such as students’ cognitive load and spatial ability were measured throughout my second educational intervention to understand how these were impacted by the introduction of AR. Furthermore, students’ attitudes to study (the affective domain) were also evaluated.

An introduction to CLT and CTML are outlined in **section 4.1**. Next, details regarding spatial visualisation and students’ attitude to study are discussed in **sections 4.2** and **4.3** respectively. As part of intervention 2, an asynchronous pretraining activity was developed, based on the concept of the Berry pseudorotation. Details of the pretraining principle, in addition to the asynchronous pretraining activity, is outlined in **section 4.4.** Educational intervention 2 served as an opportunity to collect data pertaining to students’ attitude to study, their spatial ability and cognitive load measures, in addition to their conceptual understanding of VSEPR. Descriptive statistics of these variables are presented in **section 4.5**, alongside details of the experimental design and activity development. Students’ perceptions to the educational intervention 2 and the ChemFord AR application are also discussed. The limitations of the study are considered in **section 4.6**, with concluding remarks presented in **section 4.7**.

## 4.1 Cognitive Load Theory

The importance of considering cognitive load during instruction is grounded in CLT, that posits that individuals learn best under conditions that align with cognitive architecture (Jonassen, 2009; Sweller, van Merriënboer, and Paas, 1998). This architecture is reported to comprised of a sensory register, a working memory of limited capacity, and a long-term memory of unlimited capacity (Sweller, 1988; Sweller, van Merriënboer, and Paas, 1998, 2019). Human cognition is governed by an individual’s long-term memory (Information Store Principle; Sweller, Ayres, and Kalyuga, 2011). What an individual perceives in their environment, regardless of familiarity, and how they solve complex problems are heavily influenced by immense information stores in long-term memory. However, the learning and processing of new information requires working memory resources, whose limited capacity dictates the amount of information that can be processed simultaneously (Cowan, 2001), as well as the time during which information can be retained (Vergauwe et al., 2014). In contrast, the contents of long-term memory are sophisticated structures known as *schemas* (Sweller, 1988). An individual acquires schemas throughout a lifetime of learning, which can be nested within other schemas. The information is organised according to how an individual uses it (Sweller, 1988). Therefore, the goal of instructional methods should be to support the construction of schemas by not overloading the capacities of working memory (Sweller, 2011; Sweller, van Merriënboer, and Paas, 2019).

In addition, the ways in which individuals experience cognitive load will be different (Kalyuga, 2007). For example, an individual who is inexperienced in an aspect of chemistry may learn better with the help of worked examples (see chapter 5) than with unguided inquiry. The expert chemist does not require this support. Pedagogically, the literature reports that individual working memory limitations can be overcome through a collected working memory effect, encouraged through collaborative learning (Janssen et al., 2010; Kirschner et al., 2009). From the perspective of CLT, Sepp et al. (2019) outlines that human gestures and movements may reduce cognitive load and foster germane processing through outsourcing information processing to other modalities. In multimedia research, this effect is known as enactment (Fiorella et al., 2017).

Instruction can impose three types of cognitive load on learners (Sweller, Van Merriënboer, and Paas, 1998; Paas, van Gog, and Sweller, 2010; Van Merriënboer and Sweller, 2005):

- *Intrinsic Cognitive Load* (ICL), determined by task complexity and learners' prior knowledge.
- *Extraneous Cognitive Load* (ECL), determined by instructional features that are not beneficial to learning.
- *Germane Cognitive Load* (GCL), determined by instructional features that are beneficial to learning.

In recent years, researchers have suggested a dual model of cognitive load that includes only ICL and ECL. This provides a broader interpretation to ICL, depending on the goals of learning and instruction (Leppink, 2017). It is important to note that this dual model does not deny the existence of GCL, and that the two models support the same guidelines for the design of educational activities (Leppink, 2017). Specific recommendations regarding instructional design show that ICL should be optimised by selecting tasks that match learners’ prior knowledge and experiences (Kalyuga, 2009). ECL should be minimised to reduce ineffective cognitive load (Kalyuga and Hanham, 2011; Paas, Renkl, and Sweller, 2003). When ICL is optimal and ECL is low, learners can engage in knowledge elaboration processes that impose GCL and facilitate learning.

Within an educational setting, one risk of students interacting with AR technology may be cognitive overload. However, the literature presents contradictory findings. Authors such as Akçayır and Akçayır (2017) demonstrate that the use of AR results in an increase in cognitive load measures, causing students’ working memory to process greater amounts of information when working on a learning task (Antonioli et al., 2014; Cheng and Tsai, 2013; Wu et al., 2012). In contrast, other authors provide evidence that AR can support the reduction of cognitive load, freeing up working memory capacities, instigating the generation of GCL (Goff et al., 2018; Santos et al., 2016; Sommerauer and Müller, 2018). Researchers from both sides reason in terms of empirically validated principles from both CLT (Sweller, 1988) and CTML (Mayer, 2005), and the handling of cognitive load throughout instruction.

For example, non-optimal AR application design may result in the *split-attention effect* (Mayer and Moreno, 1998). The split-attention effect occurs when an individual must perform additional mental integration processes due to the splitting of vital learning components. Consequently, a learner must split their attention to process the required information to construct a coherent model. Regarding cognitive load, this results in an increase in ECL, which expends limited capacity working memory resources which would otherwise have been available for essential learning processes (Mayer and Moreno, 1998). To address the split-attention effect, the presentation of information in a cohesive format within an augmented experience is known in CTML as the spatial and temporal contiguity principle (Ayres and Sweller, 2014). To address the first element, the spatial aspect, the physical distance between related information can be reduced, ensuring that extraneous cognitive processes are minimised. Further, the reduction of temporal separation decreases working memory resources consumed by maintaining learning components as mental representations before the essential integration process (Ayres and Sweller, 2014). Evidence of the positive effects of the spatial and temporal contiguity principle have been provided empirically (Ginns, 2006), and studied within different multimedia instructional scenarios (Schroeder and Cenkci, 2018).

Like CLT, CTML also characterises humans' cognitive architecture with a limited capacity working memory, but further assumes that working memory processes verbal and visual pictorial information in two separate channels (Mayer, 2005). As these channels, again, have limited capacity, instructional design should address both to maximise the amount of available mental resources for essential learning, and reduce ECL. This is further outlined by Mayer and Fiorella (2014) who outline 12 instructional techniques to reduce ECL and avoid cognitive overload situations.

Focusing on AR technology, previous works in the literature have reported to understand how AR influences cognitive load. In comparative studies where AR has been employed as a tool to provide guidance on a task, empirical data has shown an improvement in performance while reporting a reduction in cognitive load (Baumeister et al., 2017; Lampen et al., 2019; Yang et al., 2018). A minority of studies have found no difference between AR guidance and control conditions (Gross et al., 2018), with a minority of studies providing evidence for higher measures of cognitive load within the AR condition, indicating poorer performance outcomes (Deshpande and Kim, 2018; Friemert et al., 2019). When AR is used to assist with a task (rather than to provide direct guidance), higher performance was also observed, whilst keeping cognitive load low compared to other conditions, in about half of reported works (Buchner et al., 2022). This is evident in work such as Bellucci et al. (2018), Fischer et al. (2016), and Polvi et al. (2018). In summary, AR has been reported to compensate for the demands of visualising complex 3D representations, resulting in higher performance compared to traditional pedagogical approaches (Lai et al., 2019; Mao et al., 2017; Turan et al., 2018). However, in other works, these claims were not supported, showing that AR can lead to an increase in cognitive load, resulting in lower performance (Pu and Zhong, 2018).

## 4.2 Spatial Visualisation

Spatial ability refers to a group of cognitive functions and aptitudes that are crucial for solving problems that involve the manipulation and processing of visuo-spatial information (Carlisle, Tyson, and Nieswandt, 2015; Harle and Towns, 2011). It is one of the most widely studied domains of cognitive ability. Michael et al. (1957) states that there are two major spatial skills:

- *Spatial orientation*. A measure of the ability to remain unconfused by changes in the orientation of visual stimuli (Ekstrome et al., 1976).
- *Spatial visualisation*. A measure of the ability to mentally restructure or manipulate the components of the visual stimuli. It is characterised as a series of complicated multi-step manipulations of spatially presented information (McGee, 1979).

Bishop (1980) identified two relevant processes of visualisation: the manipulation and extrapolation of visual imagery, and the transformation of abstract relationships and non-figural data into visual terms. Visual imagery is the ability to mentally represent the visual appearance of an object. Lohman (1988), in addition to Shepard and Cooper (1982) describe a third spatial skill, *spatial relation*, which is the ability to mentally rotate an object on its axes. Spatial relation is unique, and distinct from other spatial abilities as it also involves areas associated with motor simulation in the brain (O'Shea and Moran, 2017). Spatial imagery consists of mentally representing spatial relations between the parts or locations of an object to derive an understanding to a problem. Spatial images preserve the information of an object in a form accessible to cognitive processes.

These spatial factors are mediated and supported by spatial working memory; the ability to store visual-spatial information under attentional control to complete a task (Baddeley and Lieberman, 2017). It is believed that spatial visualisation is the primary cognitive factor that influences differences in performance and is thought to have an impact on the comprehension of 3D computer visualisation (Huk, 2006; Keehner et al., 2004; Norman, 1994). Interpretations differ to the nature of these abilities and the relationships between them, with researchers proposing that they represent distinct sub-domains of spatial ability (Lohman and Kyllonen, 1983), while others suggest that visualisation is a major sub-domain (Carroll, 1993), of which orientation is merely a component.

Bodner and Guay (1997) report a highly significant correlation between spatial ability and spatially oriented tasks in general chemistry. In support of these findings, additional studies have widely recognised that spatial ability is an important contributor to the successful learning of scientific principles and academic performance (Carlisle, Tyson and Nieswandt, 2015; Carter, LaRussa, and Bodner, 1987; Sorby, Drummer, Hungwe and Charlesworth, 2005; Wai, Lubinski, and Benbow, 2009). If students have difficulty connecting observable macroscopic phenomena with the submicroscopic, how can students obtain a full appreciation of chemistry concepts (Johnstone, 1991)? To explore this link further, a 2009 study, published in the Journal of Educational Psychology, found that 45% of individuals with STEM PhDs were within the top 4% of spatial ability in a group of more than 400,000 participants (Wai, Lubinski, and Benbow, 2009). Less than 10% of individuals with a STEM PhD were below the top quartile in spatial ability during adolescence. However, at the university level, many students lack a visual vocabulary, and display difficulties visualising rotations of objects, as well as the connections between geometric structure and molecular characteristics (Tuckey, Selvaratnam, and Bradley, 1991). The transition from drawing two-dimensional (2D) constructs to imagining and manipulating the corresponding 3D object is neither natural nor easy (Guttierez, 1996).

The answer lies in developing students' visual literacy. The interpretation of symbols, in addition to understanding the particulate nature of spatial structures are essential skills that students require to solve problems in chemistry. Furthermore, chemistry education literature contains numerous reported works that demonstrate the importance of supporting students' spatial reasoning skills using molecular models (Suits and Sanger, 2013). As such, a major goal of chemistry education is to enhance students' spatial abilities to build cognitive representations of chemistry phenomena and manipulate them mentally. According to Duval (as cited in Jones, 1998), proficiency in spatial ability can be advanced by three processes:

- *Visualisation processes*. The perception of spatial relations between two objects and perceptual constancy.
- *Construction processes*. The creation of mental images and mental rotation.
- *Reasoning processes*. Solving simple problems and exercises.

Yet, spatial ability is not a skill that is taught explicitly by STEM educators and has been demonstrated to be capable of improvement over time through practice (Rahmawati, Dianhar, and Arifin, 2021; Yang, Andre, Greenbowe, and Tibell, 2003). One of the most promising affordances of immersive technology is the provision of spatial instruction. By teaching students to think in 3D, their spatial cognition can be enhanced (Moore, 1995). However, it is naive to think that the application of immersive technology in education will benefit everyone equally in relation to spatial ability (Huk, 2006; Lee, Wong, and Fung, 2010; Mayer and Sims, 1994; Wu and Shah, 2004). Insights from Hauptman (2010) report that spatial thinking, through the application of "Virtual Spaces 1.0", can be enhanced when exercised with self-regulating questions. Virtual Spaces 1.0 exercises the user's abilities build spatial images and manipulate them. The software was designed using theories of constructivism, semiotics, and component display theory.

Further studies support these findings, such as Limniou, Roberts, and Papadopoulos (2008), who used 3D molecular representations with college students to teach the reactive properties of solutions and compounds. Following 3D training, students performed better. In contrast, Urhahne, Nick, and Schanze (2009), whose comparative work examined the effects of spatial training while teaching a module on the modification of carbon, found no difference in the learning gains of experimental groups between 3D simulations and 2D images. Research into the characteristics of virtual environments have also provided further insights. Keehner and Khooshabeh (2005) found no differences between groups that had no control on rotating images and those that could freely rotate images. Ware and Rose (1999) report that colocating an interaction device and virtual image led to 35% faster response times, compared to displacing the interaction device from the virtual image. Colocation refers to the colocation of haptic and visual sensory modes. This result is supported by the work of Barrett and Hegarty (2016), where participants using a co-located interaction device displayed faster response times. Arsenault and Ware (2004) observed disrupted performance when the task involved rotation mismatches between the interaction device and virtual image. Rotation is a spatial factor that experienced chemists do instinctively, but an aspect that learners need to practice and develop over time (Stieff, 2007).

Augmented Reality (AR) can afford students the opportunity to observe molecular representations from several perspectives when rotated. Therefore, rotation is the spatial factor that will be explored throughout this chapter. A well-known and frequently used rotation test is the Purdue Spatial Visualisation Test (PSVT; Bodner and Guay, 1997), which was used for the purposes of this research (further details presented in section 4.5.1). The PSVT requires students to visualise a presented 3D object, apply a mental rotation to that object, and then select the correct new view.

## 4.3 Attitudes to Study

In addition to cognitive factors such as spatial ability, it is also relevant in research on CLT to focus on affective factors (Mayer, 2019; Paas and van Merriënboer, 2020). Students’ understanding of chemistry specific content falls under the cognitive domain, which is delivered and assessed by an extensive body of learning and teaching pedagogy. However, students’ attitudes to this content may also be congruent with higher achievement (Brown et al., 2015; Xu and Lewis, 2011; Xu, Villafane, and Lewis, 2013). Thus, considering students' attitudes and learning experiences can help to ensure quality in teaching and learning. The ideal curriculum is one that supports both gains in content knowledge and positive attitudes towards the study of chemistry. Thus, it is appropriate to measure students' attitude to chemistry study throughout their higher education (Bauer, 2008). The promotion of positive attitudes towards chemistry is an important component of chemistry higher education (Bauer, 2008; Xu and Lewis, 2011), yet the concept of attitude towards chemistry study is somewhat nebulous, often poorly articulated and not well understood.

Within this PhD project, attitude can be described as a tendency to respond to a certain chemistry stimulus, where the response has three elements:

i. A cognitive element. (What does an individual think about studying chemistry?)
ii. An affective element. (How do individuals feel about studying chemistry?)
iii. A behavioural element.

This forms a tripartite theoretical model of attitude (Rosenberg et al., 1960). The behavioural element reflects an individual's tendency to act in a particular manner regarding the stimulus. Hence, a student's attitude is of concern to an educator as it may influence the students' engagement with teaching material, collaboration with peers, and academic achievement. Other works, such as Bagozzi and Burnkrant (1979) propose that attitude can be viewed as a two-component construct comprised of cognitive and affective components. As such, these two dimensions may simultaneously account for behavioural predispositions. Thus, attitudes to chemistry may be identified and measured using the previously described structure, but only if appropriate instruments are used.

Many instruments have been developed for use in quantifying individuals' attitudes to chemistry including: the Chemistry Expectations Survey (CHEMX; Grove and Bretz, 2007), the Chemistry Attitudes and Experiences Questionnaire (CAEQ; Coll, Dalgety, and Salter, 2002), and the Colorado Learning Attitudes about Science Survey (CLASS; Adams et al., 2008). In addition, the Attitude to the Study of Chemistry Inventory (ASCI; Bauer, 2008; Brown et al., 2015; Xu and Lewis, 2011), alongside the 8-item shortened version (ASCI V2; Brandriet et al., 2011; Xu and Lewis, 2011) may also be suitable tools to quantify students' attitudes in chemistry higher education. There is also a lack of consensus regarding the methodologies that should be employed to ensure valid measures of attitudes, and that a prominent characteristic of the literature is discussion of potential problems associated with the measurement of attitude responses. This has been well-documented within science education (Gardner, 1975; Munby, 1983; Ramsden, 1997). These include issues regarding: a lack of precision over the definition of key terms, poor instrument design, failure to address issues of validity and reliability, inappropriate analysis and interpretation of data.

Although there is a complex relationship between attitude and achievement (Freedman, 1997; Steiner and Sullivan, 1984), the limited previous work on chemistry students has suggested that associations between both the cognitive and affective components of attitude, and academic performance are weak (Bauer, 2008). For example, a correlation of 0.39 between 'Intellectual Accessibility' (the cognitive domain sub-scale of the ASCI instrument) and academic performance has been reported (Bauer, 2008). Other researchers such as Xu and Lewis (2011) have reported correlations between academic achievement and both cognitive and affective components of attitude as 0.30 and 0.34, respectively. Although significant correlations between achievement and attitude have been reported in students studying chemistry at university (Brandriet et al., 2011), the strength of these associations are poor. These weak correlations between attitude and achievement may suggest that attitude is independent of, or at best, only weakly associated with achievement in chemistry higher education study. Brown et al. (2015) also reports a low correlation, suggesting that achievement is independent of students' attitudes. Within this study, higher scores for the affective and cognitive sub-scales of the instrument employed may indicate a more positive attitude to chemistry. However, this did not translate into improved academic performance.

In contrast, other previous studies have highlighted the relationship between attitude and academic achievement (Kahveci, 2015; Xu, Villafane, and Lewis, 2013). As such, a positive attitude may be congruent with higher chemistry achievement at university. The two-factor subjective test instrument utilised in this PhD project is the ASCI (V2), designed to measure a student's "intellectual accessibility". This is thought to be influenced by an individual's relevant prior chemistry knowledge (Xu, Villafane, and Lewis, 2013). Previous works have reported an interrelationship between previous chemistry academic achievement and students' intellectual accessibility and emotional satisfaction towards chemistry (Kahveci, 2015). Such findings have an important implication for educators, as students' achievement in chemistry may be improved by not only building conceptual knowledge, but by also reinforcing a positive attitude to the study of chemistry.

## 4.4 Pretraining Principle

The pretraining principle states that an individual learns more deeply from a multimedia message when they know the names and characteristics of the main concepts (Mayer and Pilegard, 2005). This lessens the cognitive load experienced when presented with novel concepts (Mayer and Pilegard, 2005). The theoretical rationale for this is that pretraining helps to guide the learner's generative processing by showing which aspects of prior knowledge to incorporate with incoming information (Moreno and Mayer, 2007). Consequently, this helps manage demands for essential processing by distributing processing to a pretraining episode that occurs before the main teaching session. The pretraining principle is closely related to the segmenting principle, which states that individuals learn better when a multimedia message is presented in user-paced segmented, rather than as a continuous unit (Mayer and Pilegard, 2005). Both principles are used in situations where processing information in a lesson could possibly overload the learner's cognitive working memory. As such, one method to prevent cognitive overload is to reduce the amount of material that a student must process, thus lowering the level of effort associated with thinking and reasoning. This prior familiarisation of new information allows a student to concentrate on understanding content material and other lesson intricacies without the cognitive overload of attempting to learn everything from scratch (Clark and Mayer, 2016). Mayer and Pilegard (2005) outline several guidelines for the successful integration of the pretraining principle which have been adopted for this research project:

- Firstly, the value of identifying the main learning concepts and important terminology during the preparation of the learning activity cannot be understated. Intrinsically, throughout the research period, I asked myself, what do learners need to know to accomplish my educational intervention, and can affordances of AR be leveraged to support this?
- Secondly, following the pretraining activity, the implementation of these terms and concepts into the main session will also be important.
- Lastly, when designing the pretraining activity, consideration must be taken to the learner's prior knowledge and relevant chemistry experience.

Previous works in the literature support the pretraining principle. Mayer, Mathias, and Wetzell (2002) applied pretraining to learning the concepts of a braking system, which depicted the possible states of each part, and described characteristics of the system. Students in the pretraining experimental group outperformed the control condition on tests of knowledge transfer and knowledge retention. In the same year Mayer et al (2002) also demonstrated that students made fewer errors ($d$ = 0.57 and $d$ = 0.75, for experiments 2 and 3 respectively), whilst also performing better on knowledge transfer tests when they received pretraining for learning geological formations. This is consistent with further studies in fields such as electronics (Kester, Kirschner, and van Merriënboer, 2006) and electrical engineering (Pollock, Chandler, and Sweller, 2002). Students who received pretraining outperformed those who did not. In addition, work has also been conducted to understand how the pretraining principle can be applied to immersive learning environments. Meyer, Omdahl, and Makransky (2019) conducted comparative work examining the learning and motivational potential of a lesson using either immersive virtual reality (iVR) technology or video. The results indicated that the use of pretraining had a positive effect on knowledge gain in the iVR condition ($d$ = 0.81). Furthermore, students reported greater levels of self-efficacy and enjoyment (Meyer, Omdahl, and Makransky, 2019).

In addition to the interaction of different media is how the pretraining is applied. Jung, Shin, and Zumbach (2021) investigated different approaches to pretraining (guided and self-directed) concentrating on their effect on cognitive load and collaborative knowledge construction within a computer-supported collaborative learning environment. The results showed that guided pretraining was more effective than self-directed pretraining in reducing measures of intrinsic and extrinsic cognitive load. There is a mutual concern on how best to identify the key concepts that should be included in pretraining, or how intensive the pretraining needs to be. Both factors were considered during the construction of my pretraining session. The concept chosen for the pretraining session was the Berry pseudorotation mechanism, a topic not covered prior to this activity. The Berry mechanism is a molecular vibration occurring in molecules of specific geometries that causes them to isomerise through the exchange of two axial ligands. (Ugi, Marquarding, Klusacek, Gillespie and Ramirez, 1971). It is the most widely accepted mechanism for pseudorotation (figure 4.1) and most commonly occurs in trigonal bipyramidal molecules, as well as molecules exhibiting a square pyramidal geometry. The pretraining activity was designed to be easily digested by students, as to not cause cognitive overload.

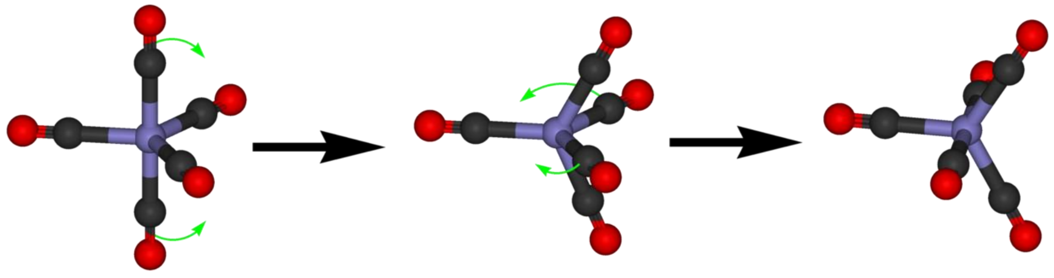


Figure 4.1. Overview of the Berry mechanism for iron pentacarbonyl

To reduce the redundancy (see section 5.1 for the expertise reversal effect; Kalyuga and Renkl, 2010) I sought to give autonomy to the learner. To support this, the pretraining (figure 4.2) was designed as an asynchronous learning activity, where students could interact with ChemFord to access an animation of the Berry mechanism. Three principles of CTML guided the pretraining design:

i. The continuity principle, which is to “align words to corresponding graphics” (Clark and Mayer, 2016).
ii. The segmenting principle, which is to break down complex information into smaller sections, which are presented sequentially (Clark and Mayer, 2016).
iii. The coherence principle, which states that all unnecessary information (extraneous material) should be eliminated (Clark and Mayer, 2016). A copy of the full pretraining exercise can be found under Appendix F.

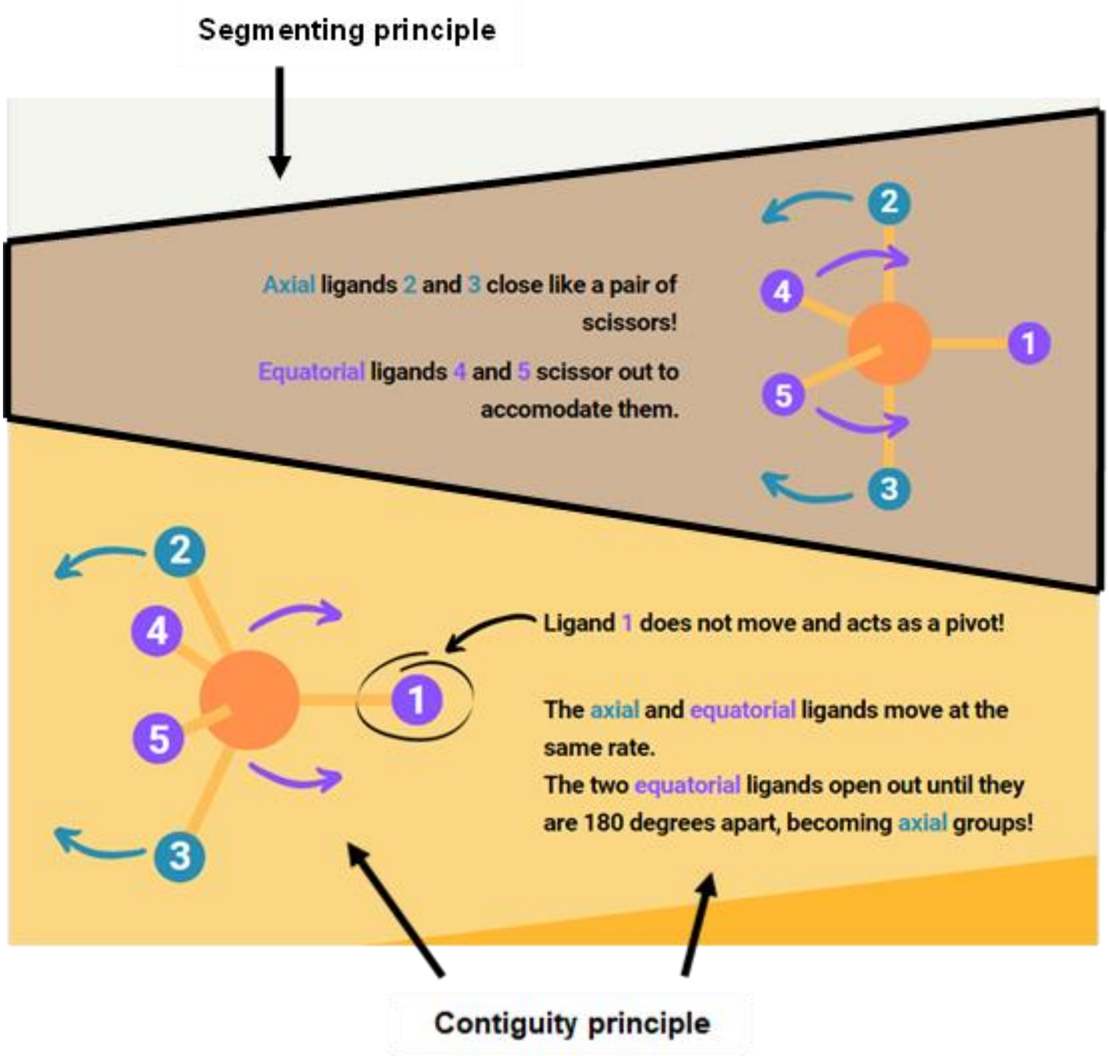



Figure 4.2. Design features of the pretraining exercise, supported by CTML principles.

Eight ChemFord virtual objects pertaining to different VSEPR geometries were developed for this educational intervention using the method outlined in section 3.3. To support the pretraining document, a Berry mechanism animation was developed and baked into the trigonal bipyramidal virtual object within Blender (figure 4.3). Students who instantiated this object could manipulate the geometry directly, in addition to toggling the Berry mechanism animation.

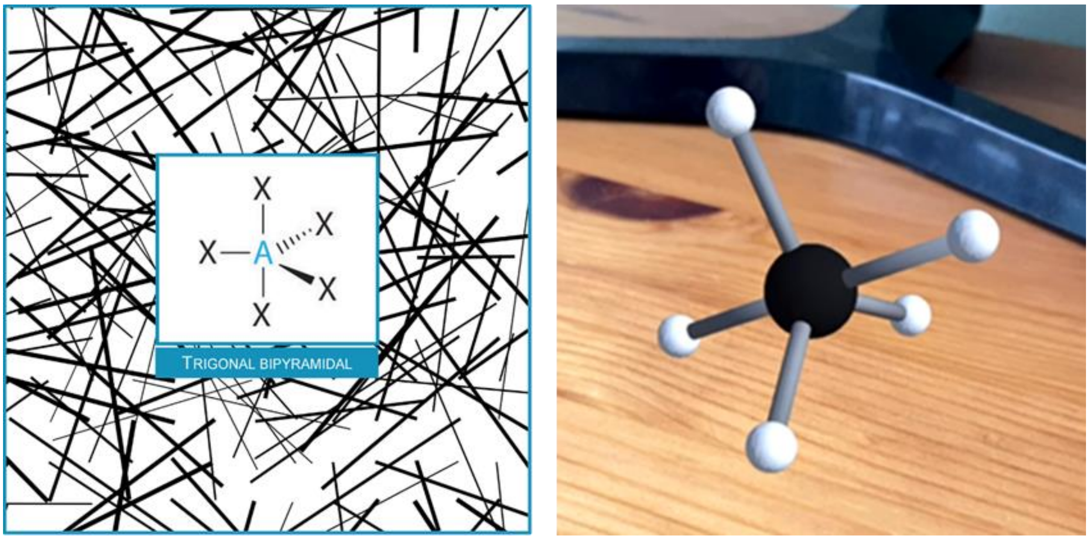



Figure 4.3. ChemFord image target for trigonal bipyramidal geometry (*left*) and trigonal bipyramidal virtual object (*right*)

## 4.5 Experimental design

A GBL activity, supported by ChemFord, was constructed to develop students' conceptual understanding of VSEPR, in addition to examining students' measures of cognitive load, spatial ability, and attitude to study. This educational intervention was conducted twice throughout the research period, following a pre-test/post-test experimental design (as outlined in figure 4.4). The participant cohort identified for this intervention were first-year undergraduate students enrolled on module "Bonding, Structure and Periodicity", the same module in which the EEA was evaluated in chapter 3. Participants were randomly assigned to one of two experimental groups to avoid bias and confounding variables:

**Experimental group 1:** The learning activity incorporated two-dimensional (2D) isometric drawings of different molecular geometries as described by VSEPR theory. This group was treated as the control group.

**Experimental group 2:** The learning activity incorporated image targets from the ChemFord which generated virtual objects of VSEPR geometries.

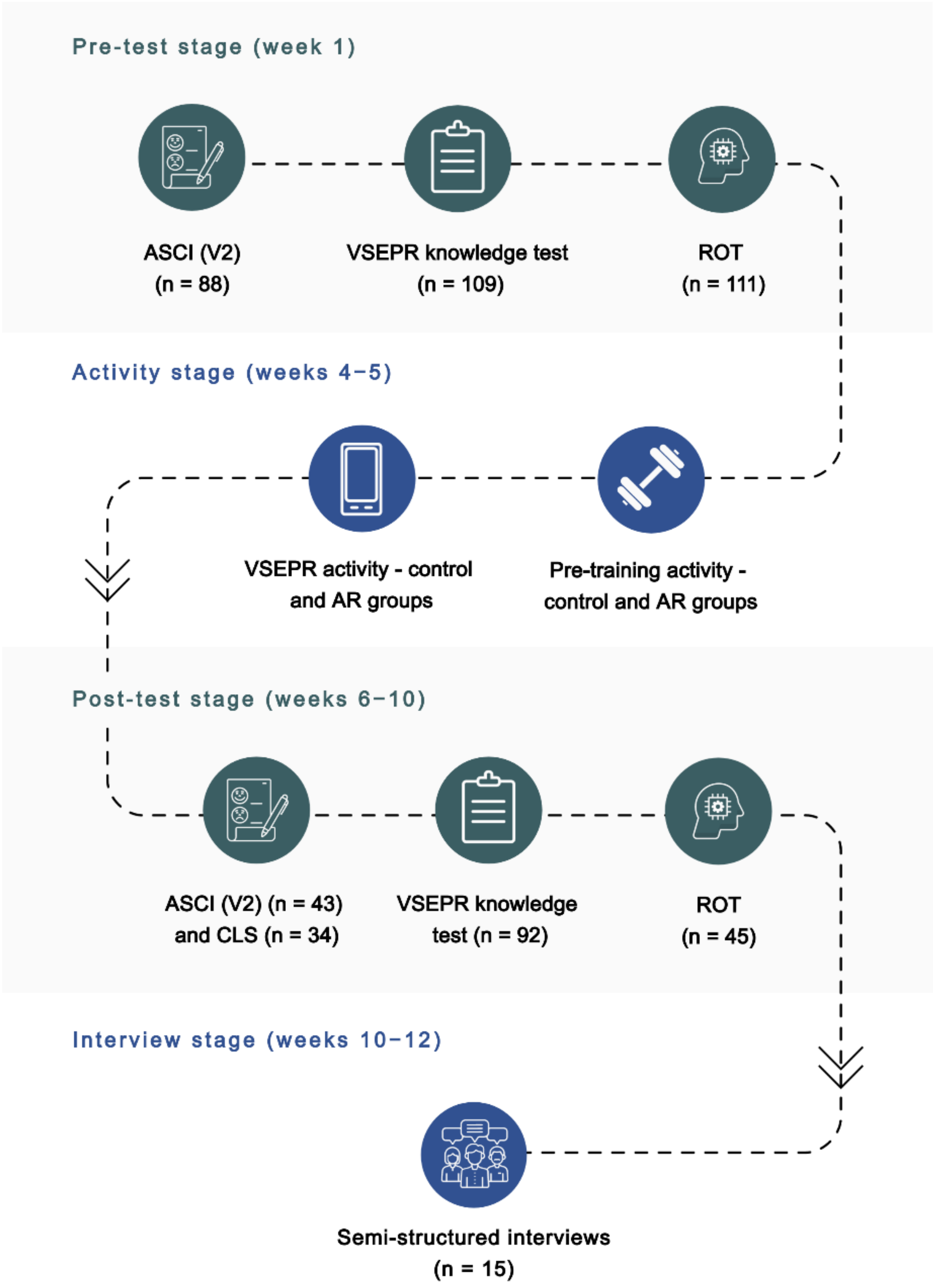



Figure 4.4. The experimental design utilised for this study, including details of participant engagement.

Throughout the two cross-sectional studies, each experimental group participated in only one version of the VSEPR intervention to eliminate carryover effects. A synchronous teaching session was conducted with the student cohort prior to the activity in which aspects of VSEPR were discussed. The research questions investigated were as follows:

**Research Question 1b.** Does the introduction of AR in an asynchronous online learning initiative improve test performance on the VSEPR test instrument of the AR group compared to the control group?

**Research Question 1c.** Does AR result in greater performance gains for students who previously exhibited lower spatial ability?

**Research Question 2b.** Do participants in the AR group display different cognitive effects for intrinsic, extraneous, and germane cognitive load compared to the control group?

**Research Question 2c.** Do students in the AR group display different responses to the Attitude to the Study of Chemistry Inventory (ASCI) compared to the control group?

**Research Question 3b.** What are the students' perceptions of the implementation of the AR technology, and my asynchronous online VSEPR learning intervention?

### 4.5.1 Test Instruments

Throughout the study period, different test instruments, taken from the literature, were employed to measure the research variables. Below, I outline details regarding each of the instruments employed.

**Cognitive Load Scale.** Students' measures of cognitive load were captured via an adapted version of the Cognitive Load Scale (CLS; Leppink, Paas, Van der Vleuten, Van Gog and Van Merriënboer, 2013). The CLS is a previously validated three-component psychometric instrument considered capable of distinguishing between ICL, ECL, and GCL (Hadie and Yusoff, 2016). This scale develops upon previous unidimensional tools that measure total cognitive load such as Paas's (1992) nine-point scale, helping researchers to determine the efficacy of learning environments as a function of instructional format and learner characteristics. This scale was adapted to the context of the VSEPR learning activity (table 4.1).

Table 4.1. The original CLS instrument (Leppink et al., 2013), alongside the adapted items used in this research study.

| # | Original Item | # | Adapted Item |
|---|---|---|---|
| 1 | The topic/topics covered in the activity was/were very complex | 1 | The topic/topics covered in the activity was/were very complex |
| 2 | The activity covered formulas that I perceived as very complex. | 2 | The activity covered molecular representations that I perceived as very complex |
| 3 | The activity covered concepts and definitions that I perceived as very complex. | 3 | The activity covered VSEPR concepts and definitions that I perceived as very complex |
| 4 | The instructions and/or explanations during the activity were very unclear. | 4 | The instructions and/or explanations during the activity were very unclear |
| 5 | The instructions and/or explanations were, in terms of learning, very ineffective. | 5 | The instructions and/or explanations were, in terms of learning, very ineffective |
| 6 | The instructions and/or explanations were full of unclear language. | 6 | The instructions and/or explanations were full of unclear language |
| 7 | The activity really enhanced my understanding of the topic(s) covered. | 7 | The activity really enhanced my understanding of the topic(s) covered |
| 8 | The activity really enhanced my knowledge and understanding of statistics. | 8 | The activity really enhanced my knowledge and understanding of molecular geometry |
| 9 | The activity really enhanced my understanding of the formulas covered. | 9 | The activity really enhanced my understanding of the molecular representations covered |
| 10 | The activity really enhanced my understanding of concepts and definitions. | 10 | The activity really enhanced my understanding of VSEPR concepts and definitions |

**The Attitude to the Study of Chemistry Inventory.** Student’s attitude to the study of chemistry was measured using the ASCI (V2) developed by Xu and Lewis (2011). The ASCI (V2) is an 8-item refinement of the original 20-item semantic differential scale developed by Bauer (2008). It measures two factors:

i. Emotional Satisfaction (the affective domain).
ii. Intellectual Accessibility (the cognitive domain).

The two aspects of attitude measured by ASCI (V2) are related, though not redundant, which is supported by two-factor confirmatory factor analysis (Xu and Lewis, 2011). The validity of the two-factor correlated structure has been confirmed in subsequent studies (Sen, Yilmaz and Temel, 2016).

**VSEPR test instrument.** An 11-item multiple-choice assessment of VSEPR chemistry achievement developed by Merchant et al (2013). The instrument examines three principles of VSEPR theory: bond angles (items 1−3); molecular geometries (items 4−8); and the identification of the shapes of molecules based on their molecular formula (items 9−11). For each of the 11 items, a score of 10 is awarded for a correct response and a score of 0 for an incorrect response. Content validity was confirmed by the authors, and Cronbach's alpha measurements suggest adequate internal consistency (Merchant et al., 2013).

**Purdue Visualization of Rotations Test (PSVT).** A widely used measure of spatial ability in science education. For this study, the revised 20-question version of the PSVT was employed (Bodner and Guay, 1997). Students are allotted 10 minutes to complete the test. For each question, students are given an example of a rotation on a 3D object, which then requires the student to perform the same rotation on a different object and choose the correct result from a pool of five options (figure 4.5). The test has consistently demonstrated good reliability across several studies (Bodner and Guay, 1997; Rahmawati, Dianhar and Arifin, 2021).

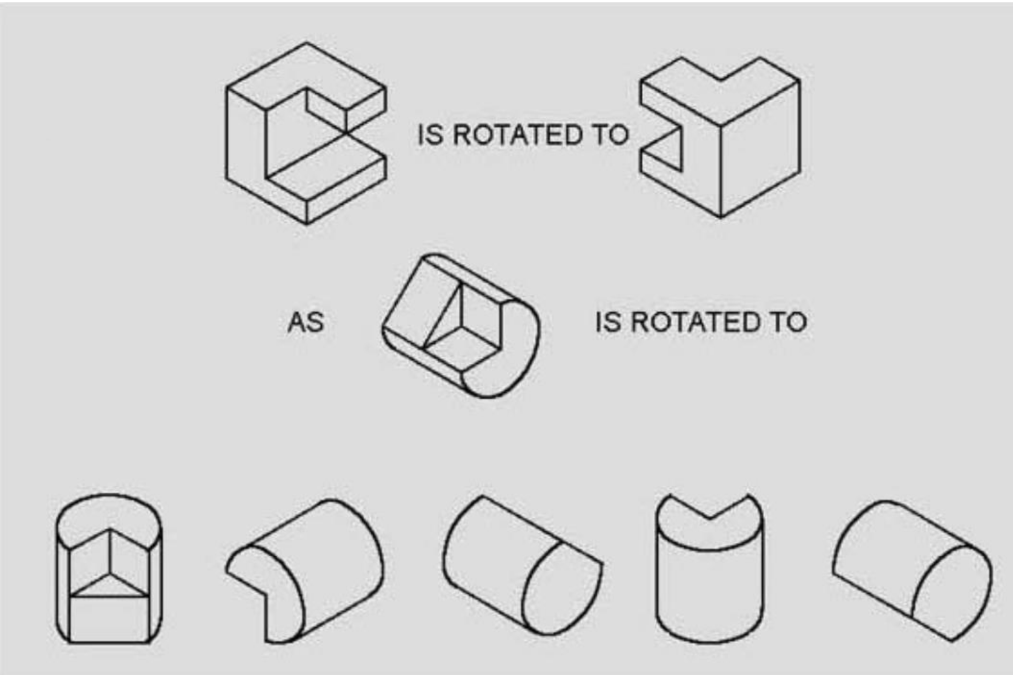



Figure 4.5. An example of an item on the PSVT.

Merchant et al. (2013) found no significant intragroup improvement in students' chemistry scores when examining students' spatial ability using the PSVT. In addition, there were no significant differences on PSVT scores between the control and experimental conditions, who utilised 2D drawings and 3D virtual models respectively. Students identified as having low spatial ability benefited from the 3D VR instruction and showed better performance than the 2D group. Furthermore, gender differences on spatial ability tests are reported in the literature. Significant differences in spatial ability, when considering gender, have been reported on several instruments including the Card Rotation Test (Ekstrom, French, Harman, and Dermen, 1976), Mental Rotations Test (Vandenberg, Kuse, and Vogler, 1985), and the Identical Blocks Test (Stafford, 1961).

### 4.5.2 Activity Design

The educational objective of this study was to develop an AR-supported GBL-based activity that could be completed synchronously or asynchronously to support students' understanding of VSEPR. The activity stage of this study was composed of two phases (figure 4.6) which were conducted in weeks 4 and 5 of the academic semester.

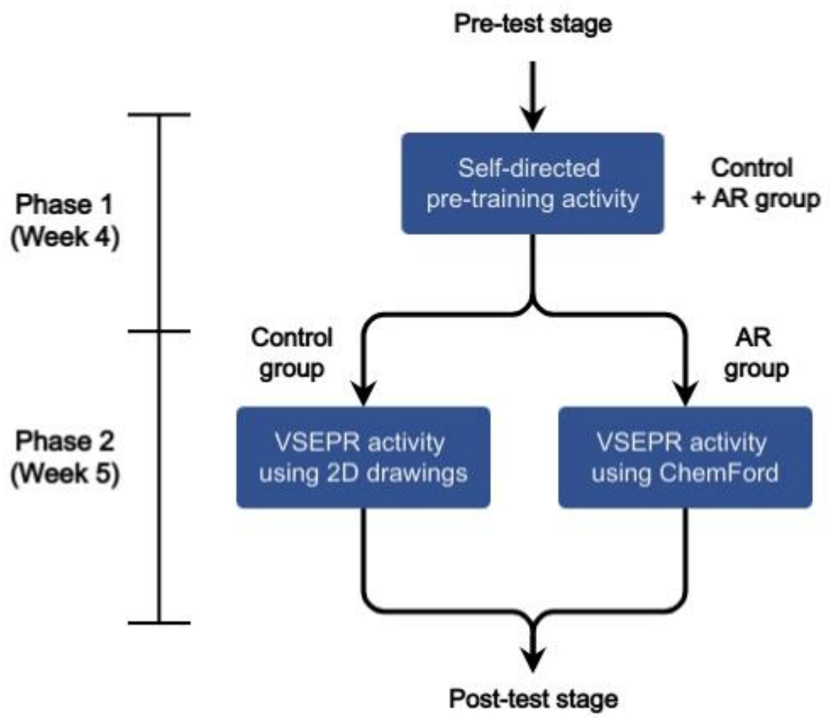



Figure 4.6. Overview of the activity stage of the study, including details of group allocation.

Phase 2 consisted of a VSEPR activity, embedding elements of GBL (Li and Tsai, 2013). In academic year 2020/2021, phase 2 was conducted asynchronously, whereas academic year 2021/2022 saw the activity conducted synchronously. This was a direct result of restrictions imposed by the COVID pandemic but presented an opportunity to see if the two approaches uncovered any significant differences. A copy of this activity can be found under Appendix G. The narrative of the activity places students as part of an expedition to the lost city of “Gillespie”. The ancient inhabitants, “Gillespians”, employed inscriptions based on molecular shapes which students must decipher to safely lead the expedition. To assist them, students are presented with the “Adventurer’s Logbook”.

The logbook provides students with a worked example of how to use the information provided by a Lewis structure to determine the correct corresponding molecular shape. Extensive research has shown that example-problem pair formats are an effective approach to problem solving (Leppink, Paas, van Gog, van der Vleuten, and van Merriënboer, 2014), particularly for the novice. The subsequent pages of the logbook outline four different collections of inscriptions, based on VSEPR theory, that students must correctly evaluate to deduce the correct path. Inscriptions for the control group were supplemented with isometric drawings of molecular geometries, whereas the inscriptions for the AR group used ChemFord image targets for generating the corresponding 3D virtual object. Addressing research question 1c, I sought to investigate the impact of the ChemFord AR visualisation aid on the ECL of the learner.

To assist conceptual understanding, overlays were developed in ChemFord to support the concept that substituents in a tetrahedral molecule are positioned at the corners of a tetrahedron, and that substituents in an octahedral molecule are positioned at the corners of an octahedron (figure 4.7). For this activity, students were required to submit their responses in long-answer format. This proved critical to evaluating whether students demonstrated a deep understanding of the topic material, and were assessed using the measurement rubric in Appendix H.

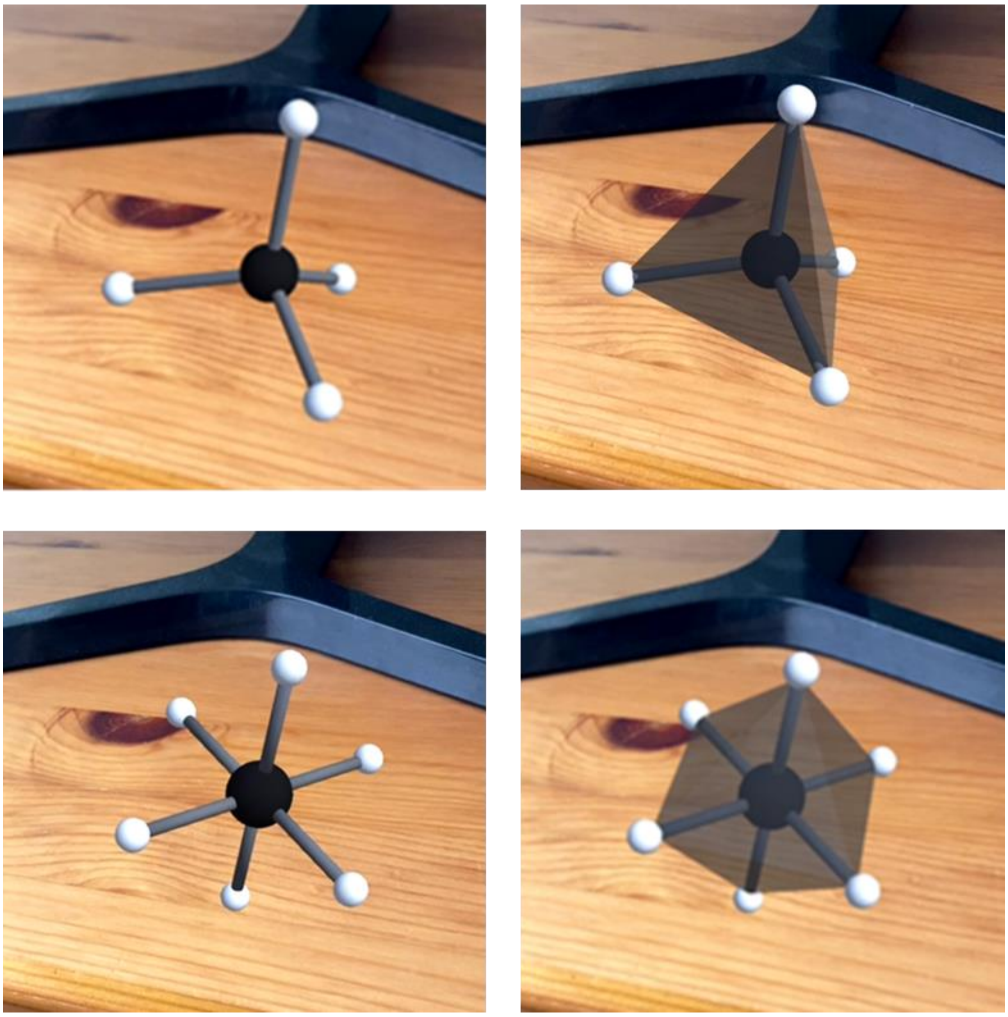


Figure 4.7. A tetrahedral geometry virtual object (top-left) with overlaid tetrahedron (top-right). An octahedral geometry virtual object (bottom-left) with overlaid octahedron (bottom-right).

### 4.5.3 Analysis of VSEPR Conceptual Knowledge Data

The descriptive statistics concerning students' conceptual understanding of VSEPR, as assessed by the VSEPR test instrument are summarised in table 4.2. The relative group-dependent means and standard deviations of the VSEPR instrument test scores obtained before and after my intervention are presented. Two sets of results are reported, from academic years 2020/2021 and 2021/2022 respectively. Following data collection in 2020/2021, I noticed that most students were answering items 9−11 incorrectly (those pertaining to the identification of the correct geometry of a molecule using its chemical formula). The average item difficulty, calculated using CTT, was 0.1. This indicates that these items were considered very difficult. As such, between the two academic years, these three questions were revised, and validated internally at UEA. Further details regarding this can be found in section 4.5.4.

Table 4.2. Relative means and standard deviations for VSEPR conceptual knowledge (0 = lowest possible score, 110 = highest possible score).

| **Academic Year** (mode of delivery) | **VSEPR concept assessed** | **Control group** Mean (SD) | **AR group** Mean (SD) |
|---|---|---|---|
| 2020/2021 (Asynchronous) | | | |
| Pre-test | Bond angles | 16.59 (10.87) | 15.83 (10.79) |
| | Geometry | 32.93 (12.70) | 33.33 (10.95) |
| | Shape determination | 2.20 (4.75) | 0.56 (2.32) |
| | **Total score** | **51.71 (21.55)** | **49.72 (18.12)** |
| Post-test | Bond angles | 23.90 (8.33) | 26.36 (6.74) |
| | Geometry | 44.63 (10.27) | 38.61 (13.34) |
| | Shape determination | 3.14 (6.17) | 2.78 (6.47) |
| | **Total score** | **71.95 (16.31)** | **65.00 (18.59)** |
| 2021/2022 (Synchronous) | | | |
| Pre-test | Bond angles | 21.82 (10.25) | 24.29 (7.87) |
| | Geometry | 40.91 (15.14) | 41.43 (6.90) |
| | Shape determination | 12.73 (10.09) | 17.14 (11.13) |
| | **Total score** | **75.45 (28.76)** | **82.86 (22.15)** |
| Post-test | Bond angles | 23.61 (8.67) | 27.14 (4.88) |
| | Geometry | 49.09 (3.02) | 48.57 (3.78) |
| | Shape determination | 22.72 (5.13) | 27.14 (4.88) |
| | **Total score** | **98.18 (13.28)** | **102.86 (7.56)** |

Across both groups, 77 students completed the instrument at both the pre- and post-test stages in academic year 2020/2021, with 18 students also completing the instrument at both stages in academic year 2021/2022. Following data collection, the Shapiro-Wilk test was used to check for the existence of normality. Although other methods for normality testing exist, Shapiro-Wilk has more power to detect the nonnormality on smaller sample sizes (Mishra et al., 2019). Data was found to be normally distributed for the pre- and post-test stages. In addition, Bartlett's test was conducted, which verified that the assumption of equal variances was true.

For the asynchronous delivery of the activity in 2020/2021, intergroup comparisons between the two experimental groups showed no significant differences in the pre-test mean scores obtained, $t(75) = 0.424$, $p = 0.666$. In addition, no significant differences were observed in the post-test mean scores achieved by the two groups, $t(75) = 1.748$, $p = 0.085$. However, it is noteworthy that significant intragroup improvements in performance between the pre- and post-test stages were observed for both the control group, $t(40) = 6.809$, $p < 0.001$, and the AR group, $t(35) = 3.884$, $p < 0.001$. This supports the premise that, for both experimental groups, chemistry instruction using a synchronous session, coupled with an asynchronous GBL activity can enhance relevant chemistry understanding. For this cohort of students, the introduction of AR technologies did not result in a significant improvement in performance on the VSEPR instrument over and above that observed for the control group.

To further investigate the post-test scores achieved by both experimental groups, an ANCOVA was performed on each of the three sections of the VSEPR test instrument. The experimental group was used as the between-subject factor, with pre-test scores as the covariate. No significant differences were found in student performance on test items pertaining to bond angles, $F(1,76) = 0.004$, $p = 0.951$, and species identification, $F(1,76) = 0.110$, $p = 0.741$. Yet, significant differences were observed for questions regarding molecular geometry, $F(1,76) = 5.508$, $p = 0.027$. Normalised change calculations were also conducted as a measure of students' learning gain between the pre- and post-test stages. The $c$ values calculated were 0.38 for the control group, and 0.26 for the AR group. To account for the variance in individual scores, the effect size was also calculated (Cohen, 2013). The effect size was calculated as 0.36.

For the synchronous delivery of the activity in 2021/2022, intergroup comparisons between the two experimental groups also showed no significant differences in the pre-test mean scores obtained, $t(16) = 0.578$, $p = 0.571$. Furthermore, no significant differences were observed in the post-test mean scores achieved by the two groups, $t(16) = 0.843$, $p = 0.412$. As observed in the asynchronous condition, significant intragroup improvements in performance between the pre- and post-test stages were seen for both the control group, $t(10) = 2.806$, $p = 0.019$, and the AR group, $t(6) = 2.763$, $p = 0.033$. Again, for this cohort of students, the introduction of AR technologies did not result in a significant improvement in performance on the VSEPR instrument when compared to the control group. An ANCOVA performed on each of the three areas assessed by the VSEPR test instrument showed no significant differences in performance on test items pertaining to bond angles, $F(1,17) = 0.009$, $p = 0.925$, molecular geometry, $F(1,17) = 0.594$, $p = 0.453$, and species identification, $F(1,17) = 0.398$, $p = 0.538$.

For comparison of the mode of delivery (asynchronous vs synchronous), only items 1−8 on the VSEPR test instrument were used. Comparison of species identification scores (items 9−11) is impossible as these items were revised following the asynchronous study (see section 4.5.4). For the control condition, no significant differences were observed on items pertaining to bond angles, $t(50) = 1.420$, $p = 0.162$, but items pertaining to molecular geometry approached significance, $t(50) = 1.778$, $p = 0.081$. No significant differences were observed between control conditions in the post-test stage for items pertaining to bond angles, $t(50) = 0.902$, $p = 0.371$, and molecular geometry, $t(50) = 1.414$, $p = 0.164$. For comparison of the AR groups, pre-test scores showed a significant difference on scores regarding bond angle items, $t(41) = 2.060$, $p = 0.046$, with students in the synchronous delivery mode scoring much higher on item 2. Pre-test molecular geometry scores approached significance in favour of the synchronous mode of delivery, $t(41) = 1.874$, $p = 0.068$. At the post-test stage, no significant differences were found on items pertaining to bond angles, $t(41) = 1.040$, $p = 0.305$, with molecular geometry items approaching significance in favour of the synchronous delivery mode, $t(41) = 1.943$, $p = 0.059$.

Normalised change calculations were also conducted as a measure of students' learning gain between the pre- and post-test stages. The $c$ values calculated were 0.64 for the control group, and 0.70 for the AR group. To account for the variance in individual scores, the effect size was also calculated (Cohen, 2013). The effect size was calculated as 0.32. The difference in normalised change scores between the synchronous and asynchronous delivery modes can be majorly attributed to improved performance on items 9−11 in academic year 2021/2022.

### 4.5.4 VSEPR Test Instrument Reliability Analysis

In their work, Merchant et al. (2013) state that content validity was conducted on the VSEPR test instrument but provide no details regarding reliability analysis. For details regarding Classical Test Theory (CTT) and Item Response Theory (IRT), see section 3.6.4. Following the initial asynchronous session in academic year 2020/2021, CTT was conducted on the 90 post-test responses of the VSEPR test instrument to understand how the items scored in terms of difficulty and discrimination (figure 4.8).

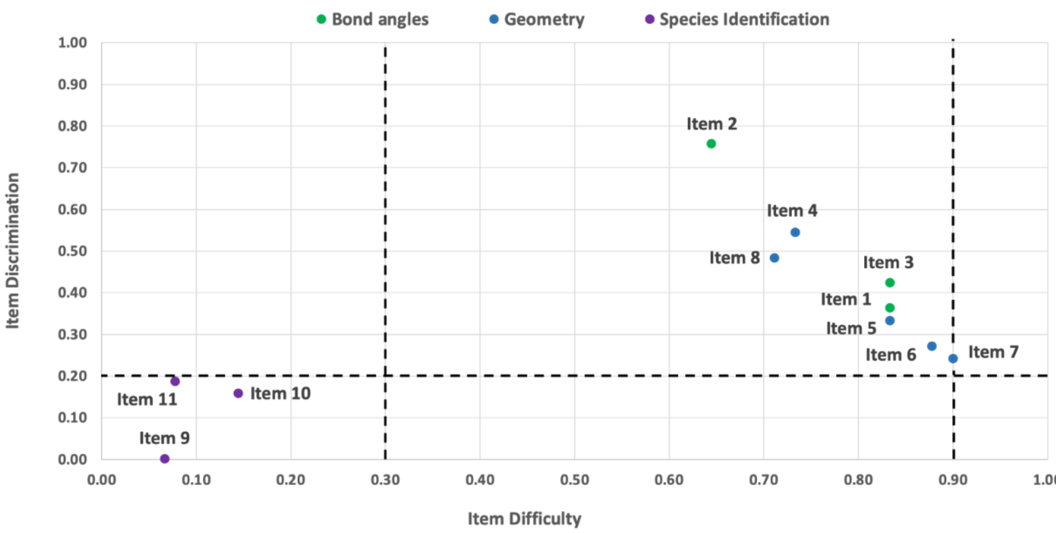



Figure 4.8. CTT analysis conducted on students' post-test responses in academic year 2020/2021. The black lines show the acceptable boundaries of difficulty and the acceptable lower bound of discrimination.

On first inspection, items pertaining to species identification scored incredibly low for values of difficulty and discrimination. As such, species identification items on the VSEPR test instrument are unable to discriminate between students with lower and higher relevant chemistry ability. However, CTT is sample dependent, and to ensure a greater level of confidence with the difficulty and discrimination values calculated, I administered the VSEPR test instrument with students on a foundation year module at the University of East Anglia (UEA) called "Introductory Chemistry" in week 8 of the academic semester. This was conducted using a pre-test/post-test design. Students received a synchronous teaching session on concepts of VSEPR prior to the post-test stage. The difficulty and discrimination values calculated, using CTT, for 60 responses from the foundation year are shown in figure 4.9.

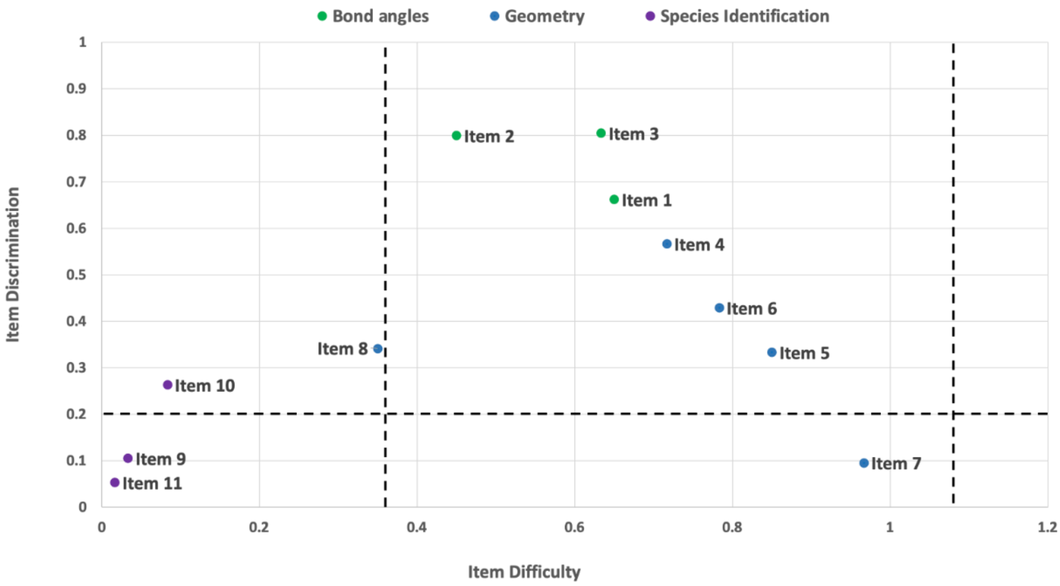



Figure 4.9. CTT analysis conducted on students on Foundation year module "Introductory Chemistry".

Once again, items pertaining to species identification scored poorly for values of difficulty and discrimination. Therefore, following the asynchronous facilitation of my VSEPR educational intervention in academic year 2020/2021, it was decided that items 9–11 would be revised prior to the synchronous cross-sectional study in academic year 2021/2022. For each of the three items, the following four changes were made:

i. Firstly, the stem of each item was rewritten.

Originally, the stem of items 9–11 read as follows:

*"You are given two 3-dimensional views of the same species. Ignore the atom colors. Pick ALL the species that has/have that shape. There may be more than one."*

For concision, and clarity, I shortened the stem of the items to the following:

*"You are given two 3-dimensional views of the same species. Pick the molecule or ion that adopts the shape represented."*

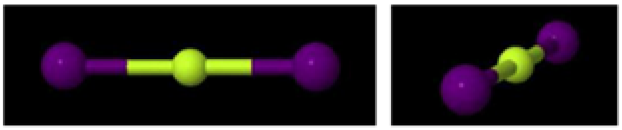
ii. Secondly, each item was presented with only one correct answer, as opposed to the multiple correct answers that students must select on the original items.

iii. Next, the number of distractors for each item was reduced from six to three.

iv. Lastly, the accompanying graphics, representing each of the molecular geometries was updated using ChemFord.

An example of these implemented changes, on item 9 of the VSEPR test instrument, is shown in figure 4.10. Items 9–11 then underwent one round of internal content validation at UEA. The results of the internal validation were positive, and no further amendments to the items were made. Upon revision, and after conducting the synchronous study of educational intervention 2 in academic year 2021/2022, CTT was again run on the revised VSEPR test instrument to calculate new values of difficulty and discrimination for each item (figure 4.11).

9.



You are given two 3-dimensional views of the same species. Ignore the atom colors. Pick ALL the species that has/have that shape. There may be more than one.
a) $H_2S$ b) $SO_2$ c) **$BeF_2$** d) **$CO_2$** e) **$BrF_2$** d) $H_2O$ e) $CaCl_2$

**Item 9:**

You are given two 3-dimensional views of the same species. Pick the molecule or ion that adopts the shape represented.

a) $H_2S$

b) $SO_2$

c) **$CO_2$**

d) $H_2O$

Figure 4.10. Item 9 on the VSEPR test instrument developed by Merchant et al. (2013) (*top*) and the revised version of item 9 for the purposes of this PhD project (*bottom*).

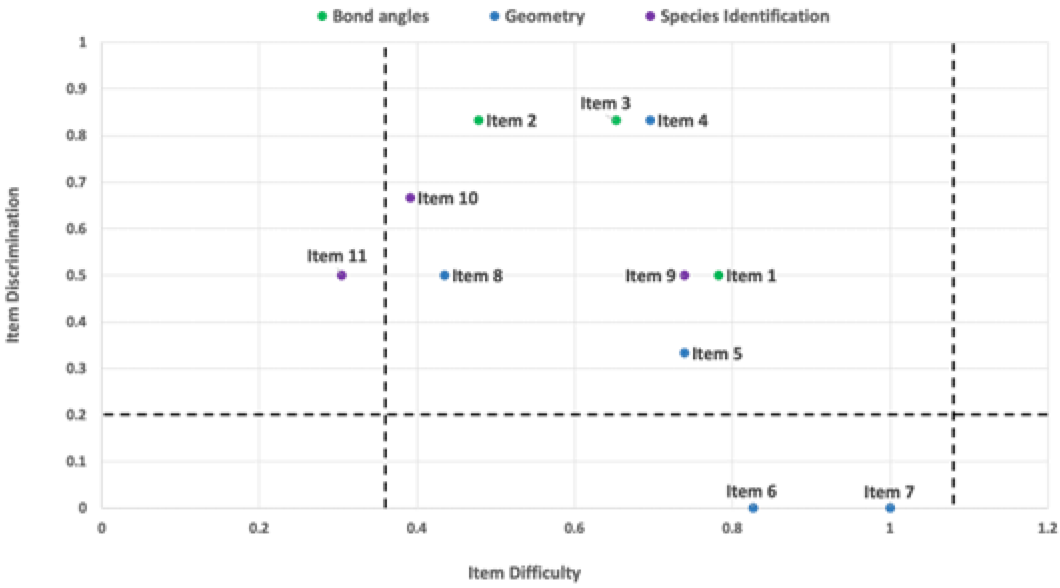



Figure 4. 11. CTT analysis on the revised VSEPR test instrument following the synchronous study in academic year 2021/2022.

The CTT analysis conducted on the 2021/2022 cohort shows acceptable values of difficulty and discrimination for items 9 and 10. Item 11 is considered the most difficult item on the instrument but displays acceptable discrimination. For this cohort, items 6 and 7 displayed low levels of discrimination. To further investigate the item properties of the VSEPR instrument, IRT analysis of the VSEPR test instrument data was performed using 3 different dichotomous 1PL, 2PL, and 3PL models. The first model assesses the first 8 items of the VSEPR test instrument. The second and third models assess the species identification questions (items 9–11) before and after revision. To assess the absolute fit of each model, two measures were examined. Firstly, a generalisation of Orlando and Thissen's (2003) $S\text{-}\chi^2$ item-fit statistic was inspected. The item-fit statistic assesses the degree of similarity between model-predicted and empirical response frequencies by item response category. A statistically significant value indicates that the model does not fit a given item (shown in blue). The $S\text{-}\chi^2$ fit statistic for each item (table 4.3) indicates a satisfactory fit in 7 of the 8 items for the 1PL model. For the 2PL model, item 1, again, displayed a non-satisfactory fit. Addition of the pseudo-guessing parameter (3PL model) introduced a non-satisfactory fit in items 3, 4, and 8.

Table 4. 3. Item-fit statistics for 1PL, 2PL, and 3PL IRT models.

| ITMC item number | 1PL | | 2PL | | 3PL | |
|---|---|---|---|---|---|---|
| | *S-χ2* | *p* | *S-χ2* | *P* | *S-χ2* | *p* |
| 1 | 18.41 | 0.003 | 12.97 | 0.011 | 16.50 | 0.001 |
| 2 | 4.90 | 0.428 | 4.54 | 0.338 | 5.61 | 0.132 |
| 3 | 6.25 | 0.283 | 5.44 | 0.245 | 7.87 | 0.049 |
| 4 | 8.70 | 0.122 | 8.98 | 0.062 | 9.58 | 0.023 |
| 5 | 1.79 | 0.878 | 1.77 | 0.778 | 4.61 | 0.203 |
| 6 | 4.47 | 0.484 | 2.36 | 0.671 | 3.54 | 0.316 |
| 7 | 2.60 | 0.761 | 1.15 | 0.887 | 1.81 | 0.613 |
| 8 | 7.65 | 0.177 | 7.18 | 0.127 | 7.84 | 0.049 |
| **Before revision** | | | | | | |
| 9 | 6.56 | 0.010 | 3.28 | NaN | 6.35 | NaN |
| 10 | 0.96 | 0.327 | 0.60 | NaN | 1.12 | NaN |
| 11 | 3.09 | 0.079 | 1.84 | NaN | 3.43 | NaN |
| **After revision** | | | | | | |
| 9 | 2.00 | 0.157 | 0.00 | NaN | 0.00 | NaN |
| 10 | 0.35 | 0.557 | 0.26 | NaN | 0.02 | NaN |
| 11 | 0.95 | 0.330 | 0.91 | NaN | 0.96 | NaN |

The fit of the 2PL and 3PL models to the data were also compared using the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) fit statistics (table 4.4). AIC estimates the quality of each model and provides a single number score that can be used to determine which of the models is the best fit for the data. AIC works by evaluating the model’s fit on the data and adding a penalty term for the complexity of the model (Akaike, 1974). A lower AIC score is better but can only be used to compare other AIC scores. The BIC is similar to AIC; however, BIC penalises the model more for its complexity, meaning that more complex models will have a worse (larger) score, and will be less likely to be significant. This is because when fitting models, it is possible to increase the likelihood through addition of parameters but doing so may result in overfitting. Overall, BIC scores are likely to be higher than AIC scores. The addition of the pseudo-guessing parameter (3PL model) did not improve the model fit. Thus, the 2PL model was used to interpret item parameters.

Table 4.4. Model level fit comparison for 2PL and 3PL models for this study.

| Model | Log-likelihood | AIC | BIC |
|---|---|---|---|
| Items 1–8 | | | |
| 2PL | −388.30 | 808.60 | 853.07 |
| 3PL | −380.60 | 809.21 | 875.90 |
| Items 9–11 (before revisions) | | | |
| 2PL | −78.70 | 169.39 | 184.39 |
| 3PL | −78.70 | 175.39 | 197.89 |
| Items 9–11 (after revisions) | | | |
| 2PL | −42.46 | 96.92 | 105.13 |
| 3PL | −42.46 | 102.92 | 115.23 |

Figure 4.12 shows the item-characteristic curves (ICC) generated from the 2PL model for items 1–8 on the VSEPR test instrument. ICCs are the fundamental unit in IRT and can be understood as the probability of answering a dichotomous item correctly, for individuals with a given ability (Goldhammer, Martens, and Lüdtke, 2017). Items that are easy to correctly answer are shifted to the left of the scale, whereas items that are difficult to answer correctly are shifted to the right. Generally, the ICC have a sigmoid curve, beginning on the left with low probabilities of answering an item correctly for lower values of student ability, rising to represent increasing probabilities of answering the item correctly as student abilities increases.

The item threshold (the point at which curve inflection occurs) indicates the item's difficulty. For the 2PL model, the inflection point occurs where the curve crosses the median probability value. This indicates the student ability for which the probability of answering the item correctly is 0.5. In addition, items also have an estimated discrimination parameter allowing item curves to have different slopes. The steeper the slope of the item response function, the better that item discriminates among students of different abilities. Students whose ability measures are at the flatter ends of the item's curve cannot be separated with a great degree of confidence by that item. Item 1 was the easiest item, generally at the lower estimate of individuals' ability. Item 10, absent from the below ICC is the hardest item. The full list of difficulty and discrimination values obtained from the 2PL IRT models are shown in Table 4.5. Figure 4.12 displays the ICC generated from my 2PL model.

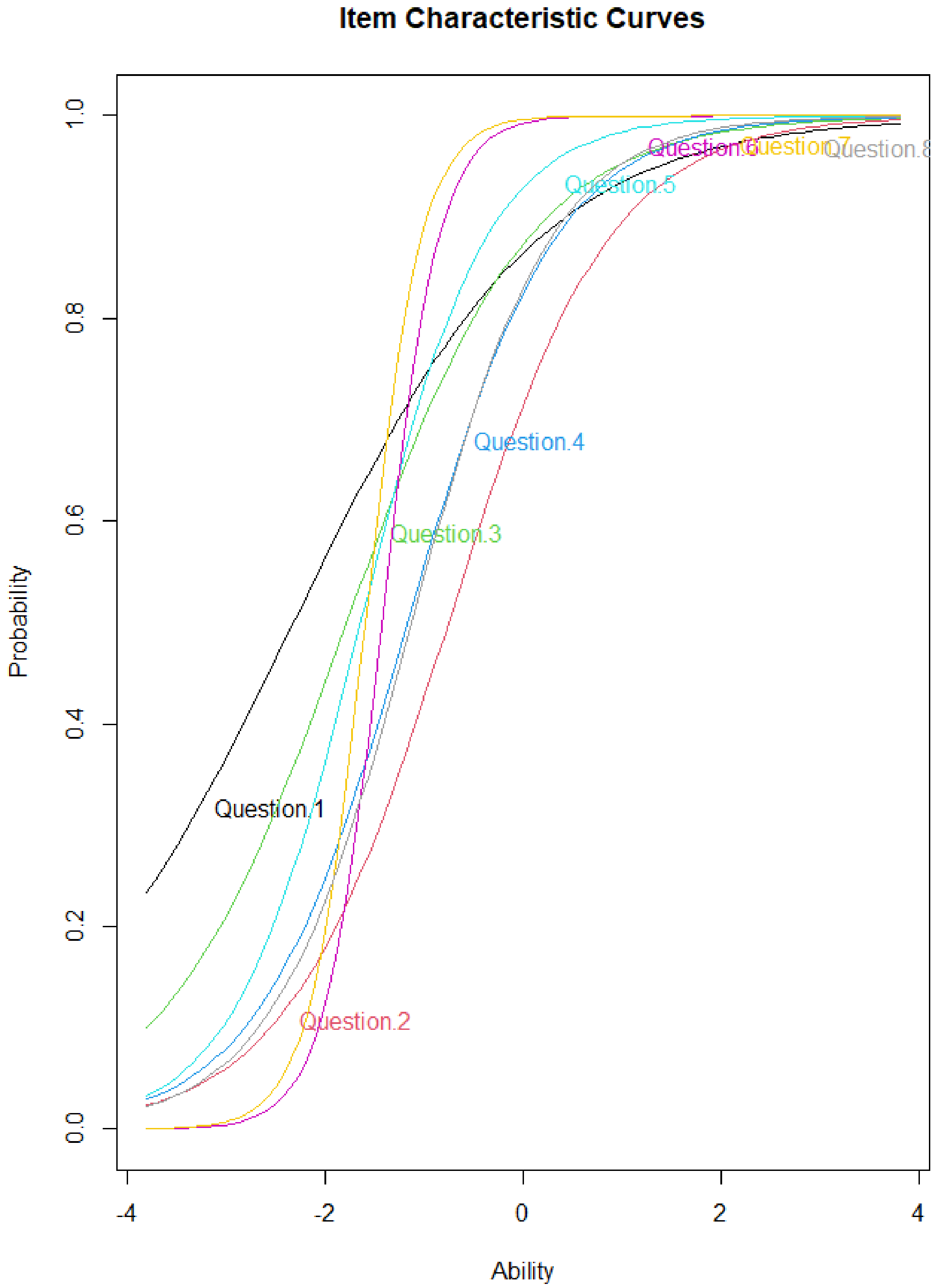



Figure 4.12. The ICC generated from the 2PL IRT model.

All items on the VSEPR instrument demonstrate good values of discrimination, with item 11 discriminating greatest between students of lower and higher ability. Following revision, items 10 and 11 displayed easier difficulty and greater discrimination, also represented by the output of CTT analysis. Item 9 also displayed easier difficulty but also demonstrated lower discrimination following revision.

Table 4.5. The difficulty and discrimination values obtained from the three conducted 2PL IRT models.

| **Item number** | **Difficulty** | **Discrimination** |
|---|---|---|
| 1 | −2.321 | 0.801 |
| 2 | −0.757 | 1.223 |
| 3 | −1.784 | 1.085 |
| 4 | −1.168 | 1.331 |
| 5 | −1.644 | 1.568 |
| 6 | −1.432 | 3.440 |
| 7 | −1.598 | 3.501 |
| 8 | −1.127 | 1.415 |
| 9 (before revision) | 1.508 | 7.918 |
| 10 (before revision) | −1.235 | −2.981 |
| 11 (before revision) | −1.854 | −2.120 |
| 9 (after revision) | −4.332 | 0.523 |
| 10 (after revision) | −0.592 | 0.363 |
| 11 (after revision) | −0.880 | 10.378 |

I employed Differential Item Functioning (DIF; Karami, 2012) to see if students of equal ability, but from different groups, have unequal probability to respond correctly to the items on the ITMC instrument (table 4.6 and figure 4.13). This is because DIF items can lead to biased measurement of ability. The DIF is stated to be uniform or non-uniform depending on whether the discrepancy in item performance between subgroups is consistent or non-consistent respectively. If an item is identified as having DIF, this is due to a source of variance not related to the structure measured by the test (Messick, 1994). DIF studies have an important role in assessing the validity of test scores (Finch and French, 2015) as the presence of DIF in the test items may reduce the validity of the test.

For this study, I employed two methods of DIF determination:

i. The Raju Signed Area Method.
ii. Lord's Chi-Squared Method.

In Raju's Area Measurement Method, the area between the ICCs of the two groups is examined to determine whether an item is DIF or not (Magis, Béland, Tuerlinckx, and De Boeck, 2010). As the area between the curves deviates away from zero, bias increases on an item (Raju, 1988). The detection thresholds used were –1.96 and 1.96, with a significance level of 0.05. In the Lord's Chi-Squared Method, the difference between the item parameter values of the two groups is tested (Magis, Béland, Tuerlinckx, and De Boeck, 2010). Variance covariance values of difficulty and discrimination parameters are examined, and the area between the ICCs of the groups is calculated (Hambleton, Swaminathan, and Rogers, 1991). The detection threshold used was 5.9915, with a significance value of 0.05. For both methods, no items were detected as DIF on the VSEPR instrument. This demonstrates that no measurement bias was present as a result of unequal probability of groups responding correctly to items on the instrument.

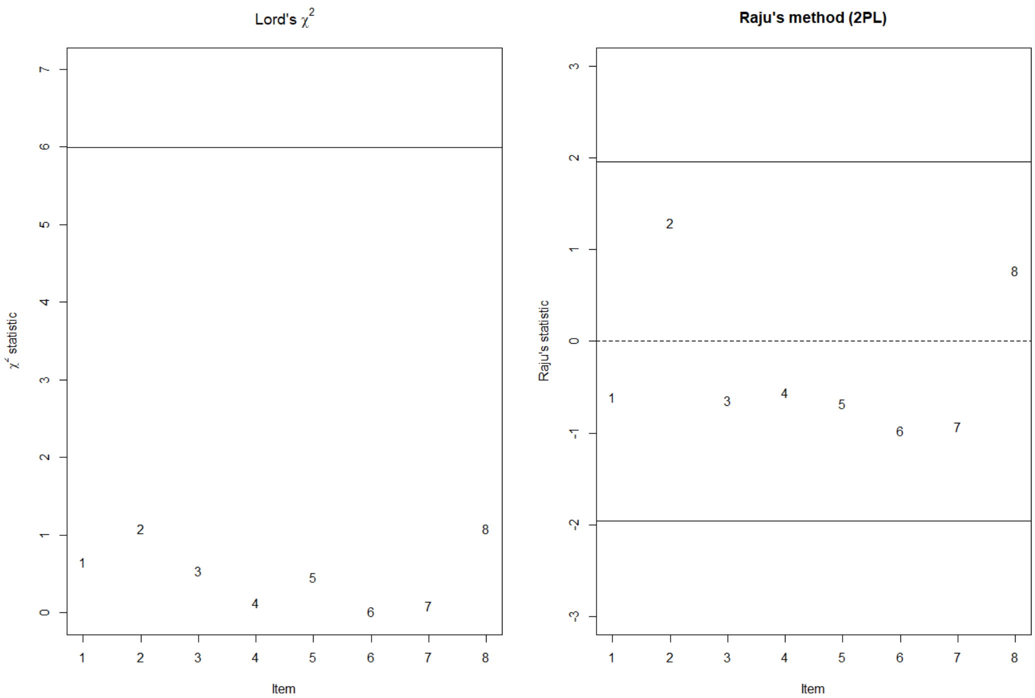



Figure 4.13. Lord (*left*) and Raju (*right*) plots on items 1−8 of the VSEPR test instrument.

Table 4.6. DIF analysis conducted on items of the VSEPR test instrument.

| Item # | Raju Signed Area method | | | Lord's Chi-square method | | |
|---|---|---|---|---|---|---|
| | Statistic | p value | DIF detected? | Statistic | *p* value | DIF detected? |
| 1 | −0.605 | 0.545 | NO DIF | 0.653 | 0.722 | NO DIF |
| 2 | 1.291 | 0.197 | NO DIF | 1.082 | 0.582 | NO DIF |
| 3 | −0.640 | 0.522 | NO DIF | 0.547 | 0.761 | NO DIF |
| 4 | −0.561 | 0.575 | NO DIF | 0.133 | 0.936 | NO DIF |
| 5 | −0.679 | 0.497 | NO DIF | 0.463 | 0.793 | NO DIF |
| 6 | −0.970 | 0.332 | NO DIF | 0.017 | 0.992 | NO DIF |
| 7 | −0.928 | 0.354 | NO DIF | 0.090 | 0.956 | NO DIF |
| 8 | 0.770 | 0.442 | NO DIF | 1.084 | 0.582 | NO DIF |
| Items 9−11 (prior to revision) | | | | | | |
| 9 | −0.000 | 1.000 | NO DIF | 0.008 | 0.996 | NO DIF |
| 10 | 0.006 | 0.995 | NO DIF | 0.053 | 0.974 | NO DIF |
| 11 | 0.009 | 0.993 | NO DIF | 0.158 | 0.924 | NO DIF |
| Items 9−11 (after revision) | | | | | | |
| 9 | −0.198 | 0.843 | NO DIF | 0.051 | 0.975 | NO DIF |
| 10 | −0.220 | 0.826 | NO DIF | 0.003 | 0.999 | NO DIF |
| 11 | −0.052 | 0.958 | NO DIF | 1.773 | 0.412 | NO DIF |

### 4.5.5 Analysis of Cognitive Load Scale Responses

Descriptive statistics concerning cognitive load measurements can be found in table 4.7, alongside measures of internal consistency (Cronbach's alpha). Internal consistency can be defined as "how closely related a set of items are as a group" (Cronbach, 1951). Technically, it is not a statistical test, but a coefficient of scale reliability. A total of 34 students completed the CLS instrument in academic year 2020/2021. However, collected responses from 2 participants were incomplete and subsequently excluded from further analysis. To reveal if significant differences for each type of cognitive load measured were present, an independent samples *t*-test was applied to each of the sub-scales. No significant differences were detected for ICL, $t(30) = 1.703$, $p = 0.099$, or ECL, $t(30) = 0.144$, $p = 0.887$. A total of 58 students completed the CLS instrument in academic year 2021/2022. Like the asynchronous condition, no significant differences were found between groups in the synchronous condition for ICL, $t(56) = 0.487$, $p = 0.628$, and ECL, $t(56) = 1.022$, $p = 0.311$.

This demonstrates that students perceived that they needed to invest similar levels of cognitive effort to understand VSEPR topic content (ICL), but also to comprehend representations of the molecular shapes (ECL), regardless of whether this was done using ChemFord or isometric drawings. For ICL, this finding is expected, and in line with the meta study by Ibáñez and Delgado-Kloos (2018). The ICL, which describes the complexity of a learning topic itself, should not be influenced by any kind of learning support such as the integration of AR technology.

I hypothesised that the introduction of AR would result in a reduction of ECL as students would exert lower levels of cognitive effort to comprehend the molecular representations. Although I did not see this result throughout this study, qualitative data collected from students may offer an insight into why this was the case (section 4.5.8). Participant interviews suggest that some of the GBL mechanics embedded into the activity required significantly higher mental effort to overcome relative to the chemistry concepts within the problems. Turan et al. (2016) report that gamification elements occupy the working memory capacities of students, therefore demanding more mental effort. This may have contributed to the ECL of the students, offsetting the cognitive advantages provided by the AR technology.

Table 4.7. Relative means and standard deviations for CLS measures (11-point scale).

| **Type of Cognitive Load** | **Control group** Mean (SD) | **AR group** Mean (SD) | **α** |
|---|---|---|---|
| Asynchronous delivery (2020/2021) | | | |
| ICL | 4.36 (2.11) | 5.53 (2.09) | 0.881 |
| ECL | 4.26 (2.37) | 4.17 (2.28) | 0.703 |
| GCL | 7.02 (2.46) | 6.50 (2.07) | 0.971 |
| Synchronous delivery (2021/2022) | | | |
| ICL | 4.33 (1.81) | 4.08 (2.09) | 0.934 |
| ECL | 3.31 (1.69) | 3.79 (1.82) | 0.737 |
| GCL | 6.84 (1.88) | 6.40 (1.97) | 0.952 |

Furthermore, for the asynchronous delivery of the GBL activity, no significant differences were observed between groups for GCL, $t(30) = 0.667$, $p = 0.510$. No significant differences were found between groups for the synchronous delivery of the activity, $t(56) = 0.853$, $p = 0.397$. This is reflected in the non-significant difference in mean scores obtained by students on the VSEPR test instrument, in line with the suggestion that GCL is indicative of information retention (Leppink, Paas, Van der Vleuten, Van Gog and Van Merriënboer, 2013). No significant between-groups effect was observed for ANCOVA results, in the asynchronous condition, when comparing ICL, with pre-test VSEPR test scores as a covariate, $F(1,29) = 2.721$, $p = 0.103$. No significant between-groups effect was observed for VSEPR post-test scores obtained with GCL as a covariate, $F(1,29) = 1.799$, $p = 0.190$. The same conclusions were drawn from the synchronous delivery of the activity. No significant between-groups effect was observed for ANCOVA results, when comparing ICL, with pre-test VSEPR test scores as a covariate, $F(1,24) = 0.319$, $p = 0.576$. No significant between-groups effect was observed for VSEPR post-test scores obtained with GCL as a covariate, $F(1,24) = 1.162$, $p = 0.293$.

For comparison of the mode of delivery (asynchronous vs synchronous), no significant differences were found in the control conditions for ICL, $t(44) = 0.432$, $p = 0.668$, ECL, $t(44) = 1.655$, $p = 0.104$, and GCL, $t(44) = 0.042$, $p = 0.967$. The same outcome was observed in the synchronous condition for ECL, $t(44) = 0.708$, $p = 0.483$, and GCL, $t(44) = 0.165$, $p = 0.870$. Significant differences were found between the ICL measures for the AR group, $t(44) = 2.367$, $p = 0.022$. However, this did not translate into greater measures of GCL.

### 4.5.6 Analysis of Spatial Ability Scores

Prior to data analysis, tests for assumptions of normality and homogeneity of variances were conducted for spatial ability data collected during both the pre- and post-test stages of academic years 2020/2021 and 2021/2022. To test normality, the Shapiro-Wilk test was employed. The collected data displayed a non-normal distribution at the pre-, $p = 0.029$, and post-test, $p = 0.003$ stages. Resultingly, Levene’s test was conducted to verify the homogeneity of variances (rather than Bartlett’s test) as this is more appropriate for non-normal data distributions. Levene’s test showed that the variances were equal at the pre-test stage, $F(2,155) = 1.559$, $p = 0.214$, but unequal at the post-test stage, $F(2,49) = 5.339$, $p = 0.025$.

A total of 51 students completed both the pre- and post-test spatial assessment throughout both academic years. The internal consistency of the PSVT was measured using Cronbach's alpha with calculated values of 0.787 for the pre-test data and 0.787 for the post-test data. This suggests good internal consistency. Due to the outcome of data assumption testing, non-parametric tests were employed for data analysis. Intergroup comparisons were conducted using the Mann-Whitney U test. For the pre-test scores achieved, no significant differences were observed between the two groups, $p = 0.880$. In addition, no significant differences were observed when comparing gender, $p = 0.409$. However, it should be noted that when comparing gender on all pre-test scores collected across both academic years (n = 157), a significant difference for gender was observed, with males performing better than females, $p = 0.005$. This result is consistent with meta-analysis conducted regarding the correlation of spatial ability and educational performance (Roach et al., 2021).

Furthermore, no significant differences were observed when comparing the post-test scores achieved by the two groups, $p = 0.850$. However, a significant difference when comparing gender at the post-test stage was observed, $p = 0.025$, with males outperforming females. Intragroup evaluations were conducted using the Wilcoxon signed-rank test. Significant intragroup improvements for spatial ability were observed in both the control, ($Z = 3.751$, $p < 0.01$), and AR group, ($Z = 4.416$, $p < 0.01$) throughout the research period. Spearman's correlation revealed a 'moderate' correlation ($r_s = 0.502$) between students' mental rotation ability, and VSEPR test scores, significant at the $p = 0.01$ level. A one-way ANCOVA showed no significant differences between-group effects for VSEPR test performance on bond angle determination, $F(1,45) = 1.030$, $p = 0.316$, and species identification, $F(1,45) = 0.031$, $p = 0.861$, when using spatial ability as a covariate. A significant difference was found for items pertaining to recognising molecular geometries, $F(1,45) = 7.727$, $p = 0.008$ when using spatial ability as a covariate, with the control group performing better.

To understand if students with lower spatial ability, who utilised AR, demonstrated greater gains in performance, a Spearman's correlation was conducted between the pre-test spatial scores obtained by students, and their calculated normalised change (figure 4.14). This was preferred over other common approaches such as the 'median split' to avoid the problems associated with categorising continuous variables (Irwin and McClelland, 2003). No significant relationship was present for this study, ($r_s = 0.214$, $p = 0.224$) and therefore further investigation of spatial ability as a predictor of performance gain, through techniques such as regression analysis, was not possible.

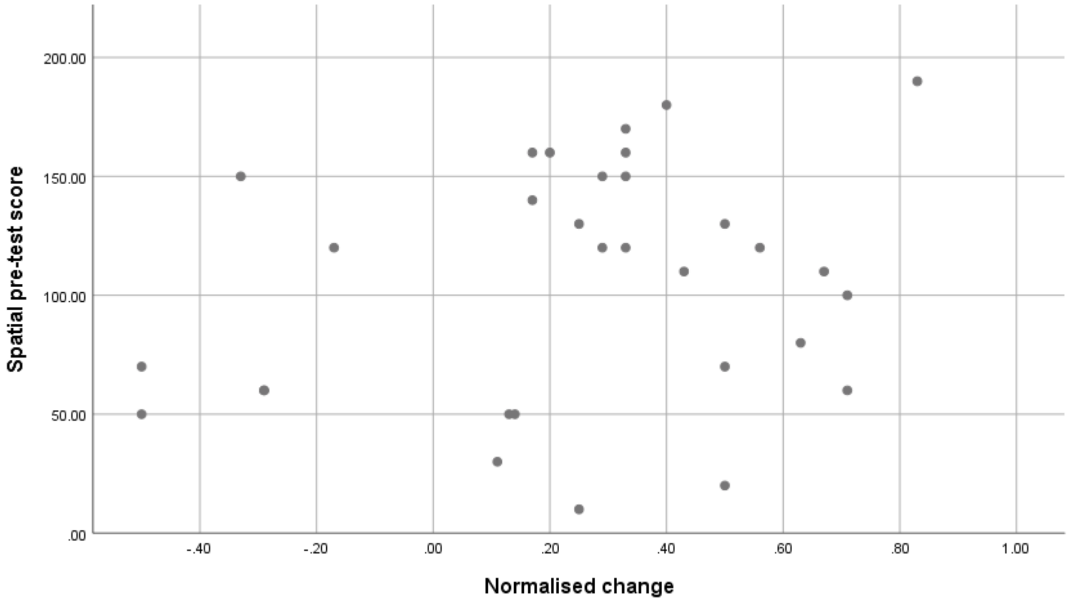



Figure 4.14. Scatter plot of Spearman's correlation examining the relationship between students' spatial ability and their calculated normalised change in the AR group.

### 4.5.7 Analysis of Students' Attitudes

To understand if students' attitudes to this intervention are congruent with higher learning gains, I am interested in exploring relationships between attitude and achievement in chemistry. The questionnaire utilised for this study, the ASCI (V2), contains two sub-scales:

i. Emotional Satisfaction (ES), corresponding to the affective domain.
ii. Intellectual Accessibility (IA), the cognitive domain.

In tables 4.8 and 4.9, each item of the ASCI (V2), as reported by both groups, is presented alongside the asymptotic significance, calculated during intergroup comparison (Mann-Whitney U test), and the group mean scores.

The internal consistency of the two sub-scales was calculated using Cronbach's alpha. For the control group, the alpha values were 0.735 (IA) and 0.767 (ES). This demonstrates good internal consistency. In addition, alpha values of 0.775 for IA and 0.735 for ES were calculated for the AR group. This indicates that a very good level of internal consistency is present. Interestingly, higher alpha-if-deleted values were calculated with regards to item 8 for both groups. A likely reason for this occurrence is the variance in meanings attributed to these adjectives, which may have resulted in students assigning different meanings to this item. Kahveci (2015) outlines the difficulty in translating item 8 to the Turkish language. It may well be that this item is not consistently interpreted by the students.

Table 4.8. Post-test ASCI responses from students in the control and AR groups of the asynchronous session.

| Item | Sub-scale | Polar adjectives | Control group Mean (SD) | AR group Mean (SD) | Asymp. Sig. |
|---|---|---|---|---|---|
| 1 R | Intellectual Accessibility (IA) | Hard/Easy | 3.73 (1.91) | 2.94 (0.93) | .247 |
| 2 | | Complicated/Simple | 4.33 (1.40) | 2.56 (0.96) | < 0.01 |
| 3 | | Confusing/Clear | 4.33 (1.44) | 4.00 (1.03) | .545 |
| 4 | | Challenging/Unchallenging | 3.27 (1.75) | 2.25 (0.77) | .358 |
| 5 R | Emotional Satisfaction (ES) | Frustrating/Satisfying | 5.27 (1.53) | 4.81 (1.47) | .086 |
| 6 R | | Uncomfortable/Comfortable | 5.00 (1.20) | 4.31 (1.20) | .066 |
| 7 R | | Unpleasant/Pleasant | 5.27 (1.33) | 4.81 (1.05) | .281 |
| 8 | | Chaotic/Organised | 4.60 (1.96) | 4.44 (1.41) | .520 |

* Items with R were reverse coded during data analysis. Items have been represented in the table in their reverse coded format.

Intergroup comparisons show no significant differences for any items from either scale, except for item 2 (Complicated–Simple). When calculating the effect size (Cohen's *d*) for item 2, a 'large' effect size of 1.542 is obtained (d > 0.8) (Cohen, 2013). I hypothesised that AR technology would simplify the visualisation of representations, however, that is not reflected in the ASCI responses collected from both participant groups. Performing a one-way ANCOVA on the VSEPR post-test scores using IA as a covariate shows no statistically significant results ($p$ = 0.488). We believe that this difference stems from discussions raised during the qualitative analysis, where students discussed the potential difficulty encountered in overcoming the gamification elements, which confounds with the potential benefits of the AR technology. However, the motivation behind the choice to study chemistry was not investigated in this study, and thus it may be possible that external factors (for example, parental and/or financial pressures) strongly influenced subject choice at university.

Table 4.9. Post-test ASCI responses from students in the control and AR groups of the synchronous session.

| Item | Sub-scale | Polar adjectives | Control group Mean (SD) | AR group Mean (SD) | Asymp. Sig. |
|---|---|---|---|---|---|
| 1 R | | Hard/Easy | 3.29 (1.27) | 3.61 (1.23) | .254 |
| 2 | Intellectual Accessibility (IA) | Complicated/Simple | 3.40 (1.38) | 3.50 (1.37) | .712 |
| 3 | | Confusing/Clear | 4.10 (1.08) | 3.93 (1.44) | .748 |
| 4 R | | Challenging/Unchallenging | 3.07 (1.20) | 2.75 (1.21) | .390 |
| 5 R | | Frustrating/Satisfying | 5.23 (1.28) | 5.04 (1.43) | .686 |
| 6 | Emotional Satisfaction (ES) | Uncomfortable/Comfortable | 4.06 (0.89) | 4.50 (1.14) | .142 |
| 7 R | | Unpleasant/Pleasant | 4.77 (1.26) | 4.75 (1.04) | .893 |
| 8 | | Chaotic/Organised | 3.74 (1.81) | 4.54 (1.82) | .094 |

* Items with R were reverse coded during data analysis. Items have been represented in the table in their reverse coded format.

For the synchronous group, the internal consistency of the two sub-scales was calculated using Cronbach's alpha. The alpha values were 0.731 (IA) and 0.526 (ES). This demonstrates good internal consistency. Intergroup comparisons show no significant differences for any items from either scale. Again, performing a one-way ANCOVA on the VSEPR post-test scores using IA as a covariate shows no statistically significant results ($p$ = 0.243). Significant intragroup differences (calculated using Wilcoxon Signed-Rank test) were found on items 1 ($p$ = 0.024), 5 ($p$ = 0.002), 6 ($p$ = 0.026), and 7 ($p$ = 0.001) for the control group. Following the activity students rated chemistry as more satisfying, more comfortable, and more pleasant, but also as slightly harder. For the AR group, significant intragroup differences were found on items 5 ($p$ = 0.003), 6 ($p$ = 0.014), 7 ($p$ = 0.002), and 8 ($p$ = 0.005). Following the activity, students rated chemistry as more satisfying, more comfortable, more pleasant, and more organised. No differences were found for items pertaining to intellectual accessibility.

Between the control groups for the asynchronous (2020/2021) and synchronous (2021/2022) cohorts, significant intergroup differences were observed for items 2 ($p$ = 0.048, Cohen's $d$ = 0.61) and 6 ($p$ = 0.002, Cohen's $d$ = 1.00) with the asynchronous group rating chemistry as simpler and more comfortable. For the two AR groups, a significant difference was observed on item 2 ($p$ = 0.043), with the synchronous AR group rating chemistry as simpler. A medium effect size ($d$ = 0.62) was found.

### 4.5.8 Analysis of Qualitative Data

Across both experimental groups, 15 students were recruited to participate in individual semi-structured interviews. The prepared interview schedule covered three topic areas:

i. The usability of the ChemFord application (including experience and interaction, and perceived usefulness).
ii. The students' experience of my intervention (including perceived learning effectiveness, satisfaction, performance achievement and reflective thinking).
iii. The cognitive benefits of integrating augmented technologies (including comprehension of topic content, problem solving, and perceived mental effort).

Qualitative analysis of the participant interview transcripts was completed through latent thematic analysis using the approach of Braun and Clarke (2006). Data was recorded, and transcribed verbatim, prior to being subjected to analysis for commonly occurring themes. The initial broad themes were constructed based on the frequency and similarity of responses. Redundancy was eliminated, and closely related major themes were merged. Throughout this thematic analysis, I focus on 3 predominant themes found in students' discussions:

i. Supporting the learning experience.
ii. AR as an asset.
iii. The challenges of integration.

I report the use of negotiated agreement as the reliability measure for this data set to minimise subjectivity in the coding process and to reduce errors. Transcripts were coded by multiple researchers independently. Any differences between the generated codebooks were discussed, and where there was a consistent disagreement, a common approach was agreed. Again, study interviewees (SI), for purposes of pseudonymisation, are represented by a number.

The first theme identified throughout this thematic analysis is **supporting the learning experience.** Throughout the discussions, students commented that the examination of molecules within the augmented environment not only reinforced their three-dimensional understanding of the VSEPR concepts, but also helped them appreciate the three-dimensional nature of chemistry, *"I always forget that molecules are, you know, three-dimensional... [this] is a constant reminder that we have to think that these are three dimensional molecules."* (SI 11). Students perceived that the integration of AR, as an additional mode of learning into the teaching process improved their understanding of VSEPR, *"With the app, I understand it better than if I was just using paper."* (SI 12). Similarly, regarding multiple contexts for learning, *"I think ChemFord definitely allows you to see it better... it actually took me quite a long time to grasp what the 2D drawings were actually trying to show."* (SI 13). In addition, students commented on their feelings of engagement with the teaching content when AR technology is utilised:

*"With the AR, if more of the lecturers did it, I would definitely like it a bit more. It breaks up the teaching content and makes it more interesting."* (SI 14).

The ability to manipulate objects within the augmented experience (moving, rotating, scaling) was considered an important affordance of the application; *"If you had a molecule that was slightly different so maybe, a mirror molecule to a different molecule, you can always compare by twisting and turning, making it bigger and smaller... And it helps me understand the difference between different molecules in different forms."* (SI 4).

The VSEPR educational intervention was positively perceived. All interviewed participants expressed a desire to repeat this style of activity in future modules throughout their degree programme. *"I would like to see more of these. I've just really enjoyed having to challenge myself in a different way."* (SI 10). Students frequently stated that the worked example in the 'Adventurer's Logbook' assisted them in correctly identifying the geometry of molecular species within the activity. Most students suggested this recurrence should be once or twice a semester (typically a 12-week period at UEA).

Participants enjoyed the challenge presented by the GBL mechanics embedded into the activity, *"It was a really nice change to just questions and bringing that sort of logic and having to think deeper"* (SI 15). Similarly, *"It's not just the chemistry but also the analytical thinking, thinking about the statements."* (SI 2). Students additionally commented that the intervention *"made me feel a bit more confident on VSEPR."* (SI 4), and that the activity *"does help you implement the knowledge that you've learned."* (SI 7).

My online VSEPR activity was primarily designed as a group activity. With the transition to online learning in response to COVID-19, we wanted to ensure social interactivity between students, and that this activity was an opportunity for students to collaborate, *"So, we did that [the activity] together in person, and having me turn something around orient it [a molecule] to show him what I was thinking. That was when I found it most helpful".* (SI 6). Yet, A minority of students commented that they *"like doing it [the activity] independently."* (SI 12). The design of the activity allows students to utilise skills both working in a team and solving problems independently.

The second theme identified throughout analysis was **AR as an asset.** A positive opinion ran throughout most participants discussions regarding the AR technology. This positivity was found both in comments regarding the affordances of AR, and in remarks regarding the alternative resources that students purport to use. For visualisation of molecular structures, students commonly mentioned the use of Molymod molecular models (Molymod, 2021). Students stated several benefits of the AR tool over physical models. Two discussion points were convenience and availability.

*"I think the AR can work better. I would have to go out and get the Molymods, whereas I can download the app and have it in 30 seconds. That was preferable."* (SI 1).

Convenience was frequently attributed to two predominate discussion points:

i. The ability to generate augmented experiences on their personal mobile devices from a large library of structures.
ii. That these structures could be created instantaneously without the additional effort of building the molecular structure.

An attributed distinction of the Molymod physical models, was the ability to modify the molecular structure; to *"take molecules apart and build whatever you like. That's quite useful."* (SI 2). This is an affordance not currently provided by the ChemFord AR application.

Students described the user interface of the application as 'intuitive', *"It's actually very easy. Very easy to use."* (SI 3). This theme was also found throughout the previously discussed thematic analysis utilising ChemFord for visualising topics of stereochemistry. As well as these descriptions, further reports of student interaction with the tool suggest minimal frustration experienced by users, an important factor in design of a tool that will be adopted by students.

*“I think it's intuitive. I mean, it always picked up markers quickly. And, you know, just tapping around at the screen, you really quickly figure out how to do stuff on it.”* (SI 4).

When discussing the mental visualisation of structures, in relation to the topic of VSEPR, the role of AR in assisting the visualisation process was of great benefit to students, in comparison to isometric drawings. *“If I can see the molecule, that’s a lot better for me. It helps me visualise.”* (SI 5). Similarly, *“The app is good for seeing things visually. I don’t really know why I wouldn’t use the app.”* (SI 6).

Lastly the **challenges of integration** emerged as the final theme of this analysis. Several challenges regarding the integration of the ChemFord AR application, and my VSEPR intervention, ran throughout participants’ discussion. Three major themes evolved from student interviews: (i) exposure of the ChemFord application, (ii) the format of the activity, and (iii) the technological limitations. Although students’ comments regarding ChemFord suggest that it was positively perceived as an educational tool, challenges were expressed regarding integration of the application into the teaching and learning process. Outside of a synchronous learning environment, students explicitly stated reasons why they may not adopt AR technology. Primarily, easy access to the image target library was seen as an obstacle for students.

*“If I had the markers to hand, it may have prompted me to look at the shapes. Not having them to hand, I just forgot about it.”* (SI 7).

Similarly:

*“I didn't use the AR, just because I didn’t have the markers to hand.”* (SI 8).

Additional accounts from interviewees describe further reasons attributing to the lower student uptake of this AR technology outside of formal synchronous activities.

*“To be honest, once it had been mentioned in lectures you kind of forgot that it was there. So, I just use Google...”* (SI 9).

In addition:

*"I think I would have been more used to using it, but because not all of the lecturers use it. It's kind of like, I haven't been shown it that much."* (SI 4).

Participants also expressed the desire to be able to toggle the requirement to scan image targets to generate the augmented experience. As an alternative, students suggested the capability to spawn objects through import from a search function. As such, I have added this as a feature to the application.

*"You're going to need a code, if you want to use it, because if they don't have anything... I think like maybe add a search bar or something with all the molecules."* (SI 5).

*"I would like to be able to keep that molecule. So, like, if you scan it could like add it to a database on the app. And you could get back that molecule, get it back up, and without having to scan the QR code."* (SI 6).

Recurrent themes of the VSEPR activity were principally coded to:

i. Difficulty.
ii. GBL elements.
iii. Affective response.

Difficulty captured students' reports of the effort required to correctly apply the VSEPR subject content to evaluate problems. The difficulty of the activity was perceived by most students to be surmountable with a minority commenting that they would have been more satisfied with a harder challenge. *"I don't think it was too bad in terms of difficulty. I thought it was at the right level."* (SI 8).

Conversely,

*"I just wish it was a little bit harder. It was really interesting and cool, and I'd love to do more things like that."* (SI 9).

A minority of students also raised comments regarding device dependent limitations of their personal mobile device when adopting AR technology. For example, students with devices that do not meet the minimum target API requirements for AR. An important step for integration of this paradigm will be to ensure accessibility for all students whilst keeping up with the rapid pace of technological developments.

## 4.6 Limitations

Some limitations of this study must be acknowledged. Firstly, a major limitation is the relatively small sample size that the data analysis was based upon. The sample size was the result of modest enrolment compounded by participant dropout between the pre- and post-test stages. Secondly, following the adoption of online learning in response to the COVID-19 pandemic, my VSEPR activity was structured as an asynchronous study activity. Consequently, I did not have the opportunity to observe students' interactions with the AR technology when participants were completing my asynchronous VSEPR intervention. Lastly, I must acknowledge the possibly of self-selection bias from participants. Students who volunteer for interviews may be different from the rest of the population regarding their communication ability or reasoning level.

## 4.7 Chapter Conclusions

This study contributes to the growing body of evidence on how students engage with embedded AR technologies. In summary, a positive opinion of my activity, and the embedded AR technology, ran throughout most participants' discussions. Students stated that the integration of AR, as an additional mode of teaching, improved their understanding of VSEPR subject content. During the activity stage of this study, participants from the AR group scored higher on submitted answers using my measurement rubric. However, this was not reflected in the post-test. Intergroup comparisons showed no significant differences on VSEPR test instrument performance. In fact, the control group was statistically better on items pertaining to molecular geometry. Further, students from both groups scored low on species identification items. Initial CTT analysis identified items pertaining to species identification as poorly discriminating and hard in terms of difficulty.

Following the activity, responses on the attitude instrument employed during this study showed that the groups scored significantly differently on item 2 of Intellectual Accessibility (Complicated – Simple). The effect size was greater than 1 standard deviation. No further significant differences in students' responses on the attitude instrument were observed.

When discussing mental visualisation of structures, in relation to the topic of VSEPR, the role of AR in assisting the visualisation process was perceived to be of great benefit to students. However, no significant differences were detected between groups for ICL, ECL and GCL. We suspect that difficulty stemming from the game mechanics confounded with the potential benefits of the AR technology. The difficulty of the activity was perceived by most students to be appropriate with a minority commenting that they would have been more satisfied with a harder challenge.

Both groups demonstrated significant improvements in spatial ability over the study period, with no significant differences observed in terms of gender performance for the post-test scores. Again, intergroup comparisons did not show any significant differences between groups. A moderate correlation was found between spatial ability and VSEPR test instrument performance.

# 5

# AR-Supported Worked Examples

Instructional guidance, provided using worked examples, helps the fledgling chemist cope with complex information, that may be difficult to process in limited capacity working memory. For students of chemistry, such complex information can pertain to the visualisation of structural changes in molecules throughout chemical reactions. Existing resources for visualising chemical reactions are largely limited to two-dimensional (2D) drawings and static physical models. While a mechanism (figure 5.1) can be used to represent the different stages of a reaction, they lack crucial user interactivity. This can be alleviated by affordances of augmented reality (AR) technology, coupled to the pedagogical approach of worked examples. Three-dimensional (3D) structure is important as it has a crucial impact on the chemical and physical properties of molecules. Within a framework of Cognitive Load Theory (CLT), this chapter illustrates how my third educational intervention (AR-supported worked examples) may enhance learning of electrophilic aromatic substitution. The participant cohort were FHEQ level 5 undergraduate students studying a compulsory module of organic chemistry. In addition, the current achievement motivation of learners was also explored, and how this may be impacted by the provision of AR technology and worked examples.

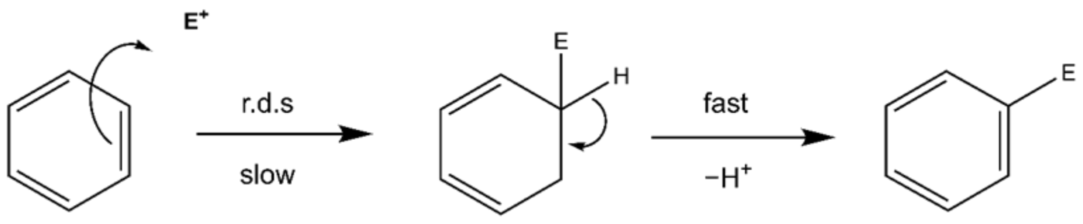



Figure 5.1. General mechanism of the electrophilic aromatic substitution reaction (Hughes-Ingold mechanistic symbol: $S_EAr$).

An introduction to the worked example effect is presented in **section 5.1**, followed by a discussion of current achievement motivation in **section 5.2**. To capture students' learning gain throughout this study, a conceptual knowledge instrument composed of items on the topic of electrophilic aromatic substitution was developed. This is referred to as the "$S_E$Ar test instrument" throughout this chapter. Details regarding the developed and content validity of the $S_E$Ar test instrument is presented in **section 5.3**. Details regarding the experimental design employed, including information pertaining to further test instruments utilised throughout the study are outlined in **section 5.4**. Within this section, the design of the worked examples, and the construction of the AR electrophilic aromatic substitution reaction is also presented. The results of the study are discussed in **section 5.5**, including a reliability analysis of the $S_E$Ar test instrument using analytical approaches of Classical Test Theory (CTT) and Item Response Theory (IRT). The limitations identified within this study are discussed in **section 5.6**, with concluding remarks presented in **section 5.7**.

## 5.1 Worked Examples

Worked examples feature regularly where problem solving is a prominent goal and are a widely studied approach to reducing cognitive load (Booth, McGinn, Young, and Barbieri, 2015; Paas, van Gog, and Sweller, 2010; Sweller, 1988). Whereas conventional problems contain only a stimulus (the description) and a stem (the problem statement), worked examples additionally outline the solution steps required to reach the correct answer. The provision of an appropriate solution reduces or eliminates random problem-solving attempts (Sweller, 2006). As such, worked examples are an empirical demonstration of the borrowing and reorganising principle.

The borrowing principle states that the knowledge held in an individual's long-term memory is borrowed from the long-term memory of others, by imitating their actions or listening to what others say, or read, or write (Chen, Woolcott, and Sweller, 2017). This process is constructive and is built upon a combination of an individual's own long-term memory, and the long-term memory of others. This results in the construction of schema that differ from both sources of information. Yet, the borrowing principle does not create new information (the randomness as genesis principle; Sweller, 2006). Throughout problem solving, new information is created using a random generation and effectiveness testing procedure (Sweller, 2006), if the information is not available in an individual's long-term memory. As described by the borrowing principle, individuals will attempt to solve a problem by using previously learned schemas. However, if the schemas are not available, one of two situations will occur. Either an individual will fail to solve a problem, or novel moves are randomly generated and tested for effectiveness (Sweller, 2006). Effective moves are retained and incorporated into long-term memory, whilst ineffective moves are discarded. Thus, the randomness as genesis principle is the only means by which new information can be obtained. However, due to the centrality of randomness in generating new information, changes must be small. Human cognitive architecture ensures this by the presence of a limited capacity working memory when dealing with new information (narrow limits of change principle; Sweller, Ayres, and Kalyuga, 2011).

Human cognitive architecture, when incorporated into CLT, can be used to predict that, for a fledgling chemist, learning via worked examples should be superior to learning via problem solving, due to the reduction of random processes. Typically, a worked example exercise is composed of two parts:

i. A worked solution to a problem with each step explained.
ii. Follow-up problems, completed by students to foster understanding of the subject content.

While the use of worked examples does not eliminate randomness, the probability of successful learning following a worked example is dramatically increased compared to learning following problem solving alone (Große, 2015). In addition, worked example study has demonstrated greater effectiveness in terms of mental effort investment (Cooper and Sweller, 1987; Hsu et al., 2015; Kalyuga et al., 2001; Mwangi and Sweller, 1998; Rourke and Sweller, 2009; Sweller and Cooper, 1985; Van Gerven, Paas, Van Merriënboer, and Schmidt, 2002; Van Gog et al., 2011). Following a worked example, learners require a procedure, normally a problem, to provide an incentive for learners to activity process the worked example, as well as to provide them with feedback on whether they have learned.

The sequence in which the two parts of a worked example exercise occur has been shown to be important. Whereas a worked example, followed by a problem, most benefits individuals with lower prior knowledge; a problem, followed by a worked example demonstrates better learning outcomes for students with higher domain-specific knowledge (Reisslein et al., 2006). This is a clear example of an expertise reversal effect (Kalyuga and Renkl, 2010). In fact, Paas and colleagues (2003) have shown in that most, if not all cognitive load effects (e.g., worked example effect, split-attention effect), reverse themselves when learners with a higher level of prior domain-specific knowledge are considered. Drawing from Snow’s (1989) aptitude-treatment interaction (ATI) theory, Kalyuga (2007) investigated the relationship between prior knowledge, cognitive load, and instructional intervention to observe the treatment effect in cognitive load. Kalyuga hypothesized that what seems difficult to low-prior-knowledge learners may prove to be easy for high-prior-knowledge learners, and vice versa (Kalyuga, 2009; Kalyuga and Renkl, 2010).

It is noteworthy at this point to introduce the concept of elements. An element is anything that needs to be, or has been learned, such as a concept or procedure (Chen, Kalyuga, and Sweller, 2015). The more elements that interact, and thus cannot be learned in isolation to achieve understanding, the greater the working memory load. This level of interaction is defined as element interactivity and is also influenced by learners’ level of expertise or prior domain-specific knowledge (Chen, Kalyuga, and Sweller, 2017). The element interactivity effect refers to the fact that if intrinsic cognitive load (ICL) is low, other cognitive load effects cannot be obtained (Sweller, 2011).

The superiority of the example-problem sequence has been demonstrated in learning materials that are higher in element interactivity (Chen, Retnowati, and Kalyuga, 2020; Sweller and Cooper, 1985). Lower element interactivity materials also allow individual elements to be learned with minimal reference to other elements, and so impose a lower working memory load. Higher elements interactivity materials consist of elements that heavily interact, and so cannot be learned in isolation. The levels of element interactivity are determined by estimating the number of interconnected elements that need to be processed at the same time to achieve understanding (Sweller and Chandler, 1994; Tindall-Ford, Chandler, and Sweller, 1997). Such estimates must simultaneously consider the nature of the information and the knowledge of the learners.

For a given task, learners with lower expertise in the task’s domain may encounter more interactive elements. Yet, multiple interacting elements for one learner with lower domain-specific knowledge may constitute a single element for a learner with a higher level of expertise. As such, learners with higher domain-specific knowledge can chunk elements to reduce the level of interactivity. With the increase in learners’ expertise, followed by the decrease in the level of element interactivity, the instructional procedures that are effective for novices may become ineffective for experts, indicating an expertise reversal effect.

For students with higher domain-specific knowledge, who have already acquired a problem schema, worked examples are no longer necessary, and may even reduce learning effectiveness (Sweller, Ayres, Kalyuga, and Chandler, 2003). In addition to the expertise reversal effect, associated implementation of worked examples should also be wary of the split-attention effect (Chandler and Sweller, 1992) and the redundancy effect (Kalyuga, Chandler, Tuovinen, and Sweller, 2001). The design of worked examples should ideally follow principles derived from CLT. According to the expertise reversal effect, as learners acquire more experience in a task domain, worked examples should be replaced with problem-solving tasks. To introduce this gradually, the use of completion tasks is suggested (Van Merriënboer, Kirschner, and Kester, 2003). A completion task provides example-style guidance for some solution steps but asks learners to complete several remaining steps on their own. A further technique involves incrementally altering worked examples by fading out solution steps. This fading technique was not found to be superior to using example-problem pairs (Renkl, Atkinson, and Große, 2004).

At the start of the learning process, a learner's low level of prior domain-specific knowledge is associated with two consequences (Kalyuga and Renkl, 2010):

i. The learner is unable to apply domain- or task-specific solution procedures. Instead, the learner must employ general problem-solving strategies.
ii. High ICL.

Within the context of worked examples, ICL is concerned with the natural complexity of any information that must be understood, not associated with instructional issues. This can only be altered by changing the nature of what is to be learned. The level of ICL for a particular task, and knowledge level, is assumed to be determined by the level of element interactivity. Any instructional issues are referred to as imposing an extraneous cognitive load (ECL). CLT is primarily concerned with techniques designed to reduce ECL. If the level of element interactivity can be reduced without altering what is learned, the load is extraneous, otherwise, the load is intrinsic. Assuming constant levels of motivation, the learner has no control over germane cognitive load (GCL, Sweller, 2011). As such, if ICL is high and ECL is low, due to organised instruction, GCL will be maximised because the learner must devote a large proportion of working memory resources to dealing with the essential learning components. GCL is slightly different in nature to ICL and ECL, in that it simply consists of working memory resources used to handle element interactivity associated with ICL.

Further, Retnowati (2018) reports that students should be reminded of any important prerequisite concepts prior to engaging with worked example problems. Throughout instruction, it is important to motivate students to understand meaningfully. Thus, for complex learning material, individually acquiring the problem and solution steps are suggested with companion of their peers. After finishing a paired problem, students may be given feedback to clarify their work. Such clarification is important to avoid students from misunderstanding, whilst satisfying students due to positive learning achievement. The authors argue that withholding instructional explanations (i.e., partial support) may provide learners with an opportunity to engage in constructive learning activities to facilitate deeper learning and far transfer, whereas materials that include full explanations could suppress inference generation because the explanatory information is already present, thereby encouraging more passive learning activities such as rehearsal and paraphrasing.

Prompting students to self-explain the rationale behind worked-out solution steps may increase GCL, if students can provide adequate explanations. However, students may lack the prior domain knowledge necessary to do so, especially very early in training. When this is the case, self-explanations are likely to induce ECL. Further, when a learner is capable of self-explanation, instructional explanations are redundant and may also impose ECL. Richey and Nokes-Malach (2013) studied partial support in worked examples and found that constructive cognitive activities were promoted which facilitated deeper understanding of the materials. There were two conditions in their study. The first condition withheld partial explanation, whilst the second condition provided a full explanation for the problems presented. The results showed that students in the withholding condition demonstrated better conceptual learning than the students in full provision condition.

## 5.2 Achievement Motivation

In addition to the cognitive load perspective is the affective perspective, which identifies relationships between learners' motivation, their cognitive load, and their prior experiences. Previous works have reported a significant correlation between GCL and measures of individuals' motivation (Um, Plass, Hayward, and Homer, 2012). As such, measures of motivation may influence the amount of cognitive resource an individual chooses to invest in a learning activity. Those learners who are self-regulated, may be able to employ more learning strategies to expand upon their effective cognitive capacity (Moreno and Park, 2010). This supports the hypothesis that higher levels of motivation can lead to greater persistence and mental effort throughout a task (Schnotz, 2010).

Within this chapter, I am interested in the concept of current achievement motivation (CAM). CAM can be defined as the instigation and aim of competence-relevant behaviour (Atkinson, 1957; Rheinberg, Vollmeyer, and Burns, 2001). In other words, why does an individual strive towards competence and away from incompetence? Rheinburg, Vollmeyer, and Burns (2001) offer a model of CAM that differentiates four distinct factors:

i. **Anxiety**. Interpreted as the fear of failure in an achievement situation.
ii. **Challenge**. Influenced by perceived task easiness and the degree to which a person accepts a task as relevant.
iii. **Interest**. The positive affect towards a task, mirroring the direct appeal the task elicits.
iv. **Probability of success**. The comparison of perceived ability and perceived difficulty. If ability outweighs task difficulty, the probability of success will be high. This factor can also be found in models of general task-motivation (Atkinson, 1957; Bandura, 1997).

Historically, a significant number of studies regarding achievement motivation have been conducted in business environments (Smith and Karaman, 2019) consisting mainly of managers and business professionals (McClelland, 1961). The findings of these studies supported the hypothesis of achievement motivation as a significant predictor of success within the business environments where the research was conducted (McClelland, 1961). In comparison, a lower volume of work has been reported on the topic of achievement motivation within an educational setting, and have produced mixed results when assessing achievement motivation as a predictor of performance (Awan, Noureen, and Naz, 2011; Kolb, 1965; Lazowski and Hulleman, 2016; Singh, 2011; Smith and Troth, 1975).

The goal of achievement-oriented tasks is to improve an individual’s capabilities in relation to a standard of competency (Heckhausen, 1977) to avoid demonstrating a lack of ability (Tanaka and Yamauchi, 2001). In this way, CAM is like self-efficacy, in that an individual’s belief in their own ability can lead to positive or negative learning outcomes. CAM is also known to be impacted by situational task characteristics; just as self-efficacy is considered an individual’s self-perception of their capabilities to accomplish a task under certain conditions (Bandura, 1977). Students will differ in their strength of motive to achieve, and educational activities will differ in the challenge that they pose. If an individual and the task characteristics display a good fit, CAM should influence task-related behaviour in a performance situation (Bipp, Steinmayr, and Spinath, 2008; Richardson and Abraham, 2009).

## 5.3 $S_E$Ar Instrument Development

In chapter 3, section 5, the development of the Isomerism in Transition Metal Compounds (ITMC) instrument was discussed as a means to assess students' conceptual understanding of stereochemistry. Again, to examine the impact of my third AR technology-supported educational intervention on conceptual understanding of electrophilic aromatic substitution, a second instrument was developed to measure students' learning gain. Throughout this chapter, the developed instrument will be referred to as the "$S_E$Ar test instrument". Again, it is important to stress that I am not trying to author a Concept Inventory (CI). However, I do believe that the PhD project benefits from the creation of an instrument containing items that can be used to quantitatively assess students' understanding of topics of electrophilic aromatic substitution.

At present, no CIs exist in the literature covering topics of electrophilic aromatic substitution. Hence, a test instrument was constructed for the purposes of this PhD project. The $S_E$Ar test instrument contains 10 items in a multiple-choice format. Responses are scored as correct or incorrect (dichotomous) which are then aggregated to yield the total score. A copy of the full test instrument can be found in Appendix I. A two-step validation approach was employed to ensure that the items on the instrument were appropriate to gauge students' conceptual understanding. Validation is the process by which a panel of experts are consulted to determine whether the instrument assesses what I intend it to assess. After the creation of the initial draft of items, internal validation with experts in the field of organic chemistry at the University of East Anglia (UEA) was carried out. I asked each consulted expert to carefully read each item, and to see whether they agreed unambiguously with the selected answer, and to comment upon whether they agreed that the item was fit for purpose. Following internal validation, one round of external validation was carried out with experts from other UK-based universities. Item changes were mostly attributed to the rewording of the stem of an item, or diagrammatic alterations. The most substantive change was made to item 1 (figures 5.2 and 5.3). The main concern was that the original version of item 1 was a lower order question (regarding Bloom's Taxonomy, 1956) which required students to remember and recall. Rather than a convergent item, item 1 was rewritten to be divergent, to encourage students to think, thus resulting in a higher order item.

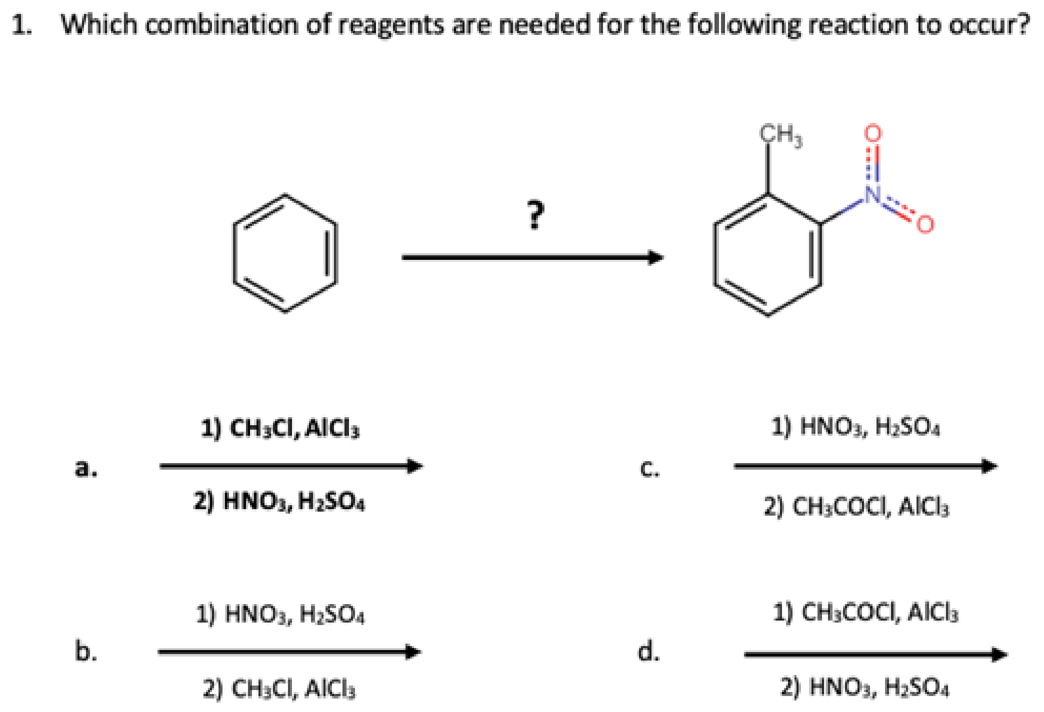
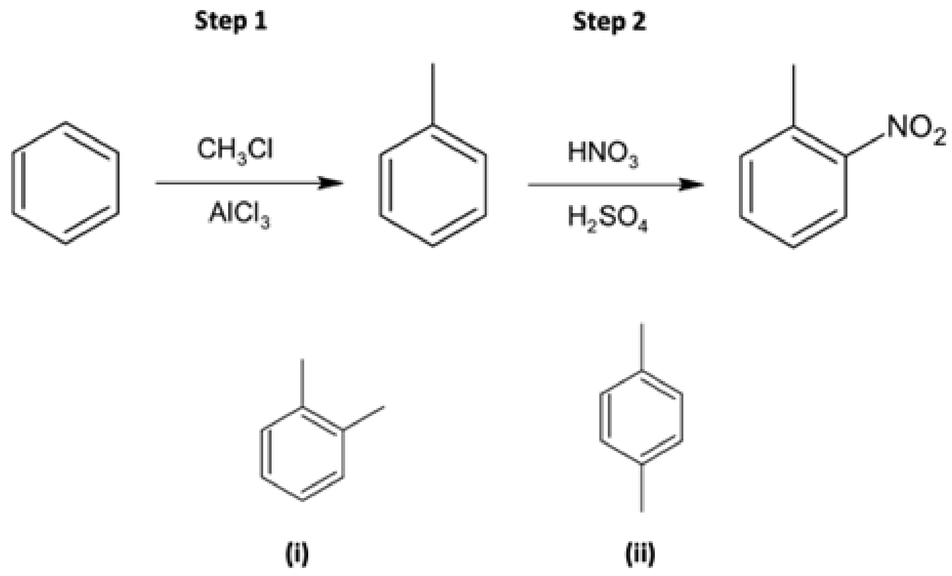



Figure 5.2. The first item on the $S_EAr$ test instrument prior to internal and external validation.

1. A student is producing nitrotoluene from benzene via the following reaction. The student observes that large quantities of compounds (i) and (ii) are produced during step 1. How can the amounts of (i) and (ii) be minimised?




a. Use a reduced amount of the $AlCl_3$ catalyst with the same amount of $CH_3Cl$

b. Carry out the reaction in step 1 at a lower temperature

c. Use an excess amount of benzene relative to $CH_3Cl$

d. Use a 1:1 ratio of benzene and $CH_3Cl$, with $FeCl_3$ as the Lewis acid catalyst

Figure 5.3. The first item on the $S_EAr$ test instrument following internal and external validation.

## 5.4 Experimental Design

The evaluation of my third educational intervention was conducted throughout academic year 2021/2022 as part of a FHEQ level 5 module of organic chemistry study at UEA. When considering the experimental design for this study, attention was given to the literature, which has criticised the use of worked examples in comparative research without using instructional support as a control condition (Koedinger and Aleven, 2007). Hence, within this chapter, I examined alternate-format worked examples couple with the same faded problems. A faded problem is one that omits steps but retains much of the guidance provided by the context of a solved example. For this study, a pre-test/post-test experimental design was employed (figure 5.4). Participants were randomly assigned to one of two experimental groups to avoid bias and confounding variables:

i. **Control group**. The worked examples incorporated two-dimensional (2D) drawings of electrophilic aromatic substitution reaction mechanisms.
ii. **AR group**. The worked examples incorporated an interactable electrophilic aromatic substitution reaction mechanism afforded by AR technology.

Each group participated in only one worked example activity to eliminate carryover effects. The pre-test stage of the study was carried out in week 2 of the academic semester and consisted of the $S_EAr$ test instrument and the Questionnaire on Current Motivation (QCM). In week 3, a synchronous teaching session was conducted with the entire student cohort prior to the activity which introduced concepts pertaining to electrophilic aromatic substitution. The activity was conducted in week 4, in which students also completed the Cognitive Load Scale (CLS). Details regarding the QCM and CLS are presented in section 5.4.1. The post-test stage in week 5. At the post-test stage, students completed the $S_EAr$ test instrument for the second time, in addition to also completing semi-structured interviews. Throughout the qualitative data collection stage, I conducted discussions with participants on topics relating to electrophilic aromatic substitution. This proved critical to evaluating whether students demonstrated a deep understanding of the topic material. Details of the interview schedule can be found in Appendix J.

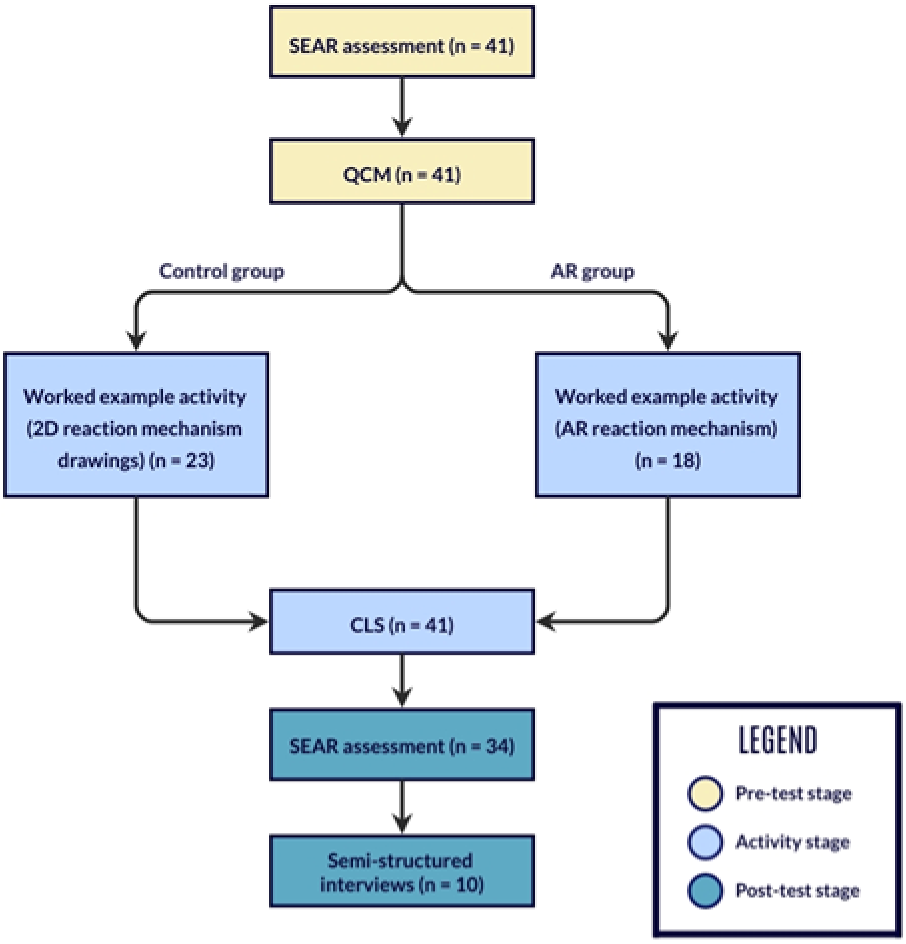



Figure 5.4. The experimental design employed, with details of participant engagement.

This study attempts to explore how coupling the pedagogical approach of worked examples with AR technology impacts students' conceptual understanding of electrophilic aromatic substitution. Further, I was interested in the interactions between students' CAM, cognitive load, and cognitive information processing. Qualitative data collection was also undertaken on students' perceptions of the learning activity, alongside a short discussion of their conceptual understanding. The research questions investigated were as follows:

**Research Question 1d**. How is relevant chemistry experience impacted by the presentation of the worked examples (AR vs 2D)?

**Research Question 1e**. Is there an expertise reversal effect signifying interactions between the mode of representation (AR vs 2D), relevant chemistry experience, cognitive load, and CAM?

**Research question 2d**. How do cognitive load measures of participants correlate with measures of CAM?

**Research Question 3c**. What are the participants' perceptions to the use of worked examples, and how do participants convey their understanding of electrophilic aromatic substitution in conversation?

### 5.4.1 Test Instruments

In addition to the $S_EAr$ test instrument described in section 5.3, the following instruments were also utilised throughout this study:

**Cognitive Load Scale.** Students' cognitive load was measured via an adapted version of the CLS (Leppink et al., 2013). The CLS is a previously validated three-component psychometric instrument considered capable of distinguishing between intrinsic cognitive load (ICL), extraneous cognitive load (ECL) and GCL (Hadie and Yusoff, 2016). This scale develops upon previous unidimensional tools that measure cognitive load such as Paas's (1992) 9-point scale, helping researchers to determine the efficacy of learning environments as a function of instructional format and learner characteristics. The scale was adapted to the context my electrophilic aromatic substitution worked examples activity (table 5.1). The CLS was distributed directly after the educational intervention had taken place.

Table 5.1. The original CLS instrument (Leppink et al., 2013), alongside the adapted items used in this research study.

| # | Original Item | # | Adapted Item |
|---|---|---|---|
| 1 | The topic/topics covered in the activity was/were very complex | 1 | The topic/topics covered in the activity was/were very complex |
| 2 | The activity covered formulas that I perceived as very complex. | 2 | The activity covered reaction mechanisms that I perceived as very complex |
| 3 | The activity covered concepts and definitions that I perceived as very complex. | 3 | The activity covered electrophilic aromatic substitution concepts and definitions that I perceived as very complex |
| 4 | The instructions and/or explanations during the activity were very unclear. | 4 | The instructions and/or explanations during the activity were very unclear |
| 5 | The instructions and/or explanations were, in terms of learning, very ineffective. | 5 | The instructions and/or explanations were, in terms of learning, very ineffective |
| 6 | The instructions and/or explanations were full of unclear language. | 6 | The instructions and/or explanations were full of unclear language |
| 7 | The activity really enhanced my understanding of the topic(s) covered. | 7 | The activity really enhanced my understanding of the topic(s) covered |
| 8 | The activity really enhanced my knowledge and understanding of statistics. | 8 | The activity really enhanced my knowledge and understanding of electrophilic aromatic substitution |
| 9 | The activity really enhanced my understanding of the formulas covered. | 9 | The activity really enhanced my understanding of the reaction mechanisms covered |
| 10 | The activity really enhanced my understanding of concepts and definitions. | 10 | The activity really enhanced my understanding of electrophilic aromatic substitution concepts and definitions |

**Questionnaire on Current Motivation**. An 18-item instrument designed to measure the four distinct factors of current achievement motivation in specific performance situations. The QCM utilises a 7-point Likert scale and has been previously shown to be a predictor of performance in a variety of complex problem-solving tasks (Freund, Kuhn, and Holling, 2011; Rheinberg, Vollmeyer, and Burns, 2001; Vollmeyer and Rheinberg, 2006). Although a short form of the instrument (consisting of 12 items) has also been developed (Freund, Kuhn, and Holling, 2011), the full 18-item instrument was employed for the purposes of this study. Previous validity and reliability analysis of the QCM has been undertaken (Vollmeyer and Rheinberg, 2006), and evidence for the absence of measurement bias on the instrument has also been provided (Freund, Kuhn, and Holling, 2011). The QCM was distributed directly before the teaching activity. The Cronbach's alpha value (Cronbach, 1951) calculated was 0.739 which shows good internal consistency. Interestingly, the removal of no items would result in a higher alpha-if-deleted value.

### 5.4.2 Activity Design

The vision for my third educational intervention draws on the coupling of worked examples with faded practice problems to elicit interaction, an approach that has been shown to yield greater learning outcomes than the use of either approach independently (Atkinson, Renkl, and Merrill, 2003; Crippen and Brooks, 2009; Jones and Fleischman, 2001; Sweller, van Merriënboer, and Paas, 1998). Progressive fading can direct students' attention to important steps (Hilbert, Renkl, Kessler, and Reiss, 2008), and allows for gradual adaptation of support in response to students' increase in conceptual knowledge, thus removing redundant information (Low, Jin, and Sweller, 2011). With reference to CAM, I am interested in the interaction of this worked example educational intervention, as the situational stimulus, with students' underlying motives. Thus, at the beginning of the session, I introduced the worked example activity and the new features of the ChemFord application. This was to ensure that students understood the cognitive demands, in addition to the requirements of the activity beforehand. Students were instructed to study the worked examples prior to attempting the faded practice problems.

The gradual fading of worked solutions in a worked example (omitted steps), which has been paired with practice problems has been previously examined (Atkinson, Renkl, and Merrill, 2003). Regarding the order in which steps can be faded:

- The final step can be the first to be omitted, with the consecutive fading of previous steps (backwards fading).
- The first step can be omitted, with the consecutive fading of subsequent steps (forward fading).

Upon investigation, Atkinson, Renkl, and Merrill (2003) found that both approaches yielded positive results, but that the backward fading technique was more time efficient. Similar results are reported by ter Vrugte et al. (2017). Students who utilised faded worked examples needed fewer attempts and less time per problem. Hence, the three faded problems in my educational intervention were increasingly backwards faded. A worked example of a backwards faded missing value problem (ter Vrugte et al., 2017) is shown in figure 5.5, where students must calculate values whilst maintaining a consistent ratio across each column. The missing information is represented as blanks. Learners are expected to gain understanding of the solution steps by filling in the blanks according to the instructions provided.

| | | | |
|---|---|---|---|
| Value 1 | 15 | 5 | ? |
| Value 2 | 6 | 2 | 18 |

÷ ×

**Worked example level 1:**

- A table
- Information about column content.
- Amounts of the first ratio (column 1)
- Partial solution 1 (column 2)
- Given amount of final ratio (column 3)
- Information about actions

| | | | |
|---|---|---|---|
| Value 1 | 18 | ... | ? |
| Value 2 | 6 | ... | 7 |

÷ ×

**Worked example level 2:**

- A table
- Information about column content.
- Amounts of the first ratio (column 1)
- ~~Partial solution 1 (column 2)~~
- Given amount of final ratio (column 3)
- Information about actions

**Worked example level 3:**

| | | | |
|---|---|---|---|
| Value 1 | 12 | ... | ? |
| Value 2 | 16 | ... | 8 |

- A table
- Information about column content.
- Amounts of the first ratio (column 1)
- ~~Partial solution 1 (column 2)~~
- ~~Given amount of final ratio (column 3)~~
- ~~Information about actions~~

**Worked example level 4:**

| | | | |
|---|---|---|---|
| Value 1 | ... | ... | ? |
| Value 2 | ... | ... | ? |

- A table
- ~~Information about column content.~~
- ~~Amounts of the first ratio (column 1)~~
- ~~Partial solution 1 (column 2)~~
- ~~Given amount of final ratio (column 3)~~
- ~~Information about actions~~

Figure 5.5. An example of a backwards faded worked example.

The design of this educational intervention draws on the principles of CLT. To maximise GCL, the resources of my worked example activity were designed to optimise ICL, that is, to be at the appropriate level of complexity. This is assumed to be determined by the level of element interactivity (Atkinson, Renkl, and Merrill, 2003). Understanding topics of electrophilic aromatic substitution requires underlying knowledge of principles of organic chemistry, and can therefore be assumed to be of higher element interactivity. When the combined ICL and ECL exceed working memory capacity, learning can be inhibited. Thus, it is essential to design instruction in a format that reduces working memory load to manageable proportions. To accomplish this, principles of the Cognitive Theory of Multimedia Learning (CTML; Mayer and Fiorella, 2014) guided the design of the learning material. I aim to minimise ECL by eliminating split attention and redundancy conditions wherever possible.

My learning resource is composed of seven sections. Firstly, drawing on the segmenting principle (Clark and Mayer, 2016) sequential chunks of information on the fundamental aspects of electrophilic aromatic substitution are provided. A full copy of the worked examples activity resource can be found in Appendix K. These include:

- Directing effects (shown in figure 5.6).
- Activating and deactivating groups.
- Regioselectivity.

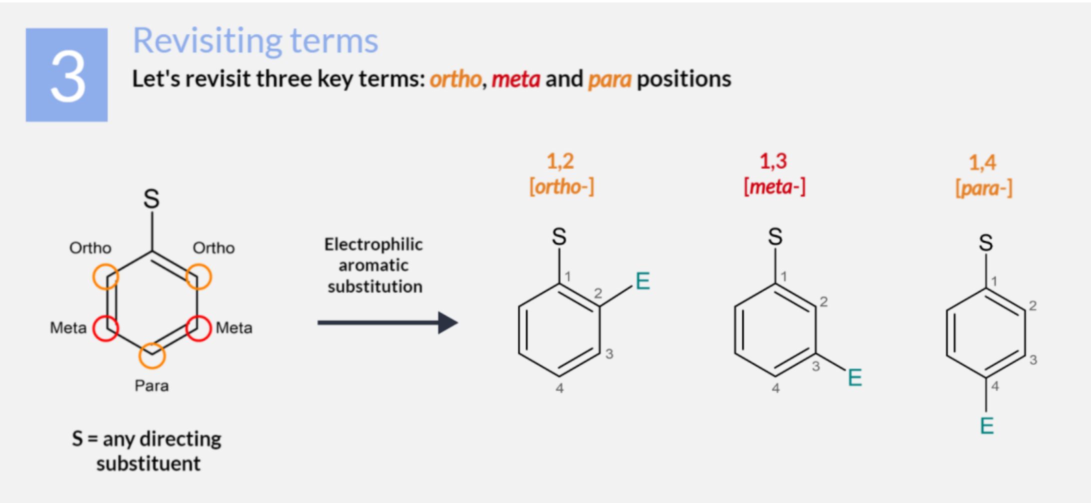



Figure 5.6. The third section of the worked example activity, displaying directing effects.

Based on learner characteristics, and the representation of educational content, previous research has reported that diagrams are more effective than textual representations (Ainsworth and Loizou, 2003). Hence, the representation of the worked examples was mainly graphical. In addition, previous work has also found positive impacts for textual explanations (Atkinson et al., 2000). With consideration to the spatial and temporal contiguity principle, words and pictures should be presented simultaneously, and near one another (figure 5.7).

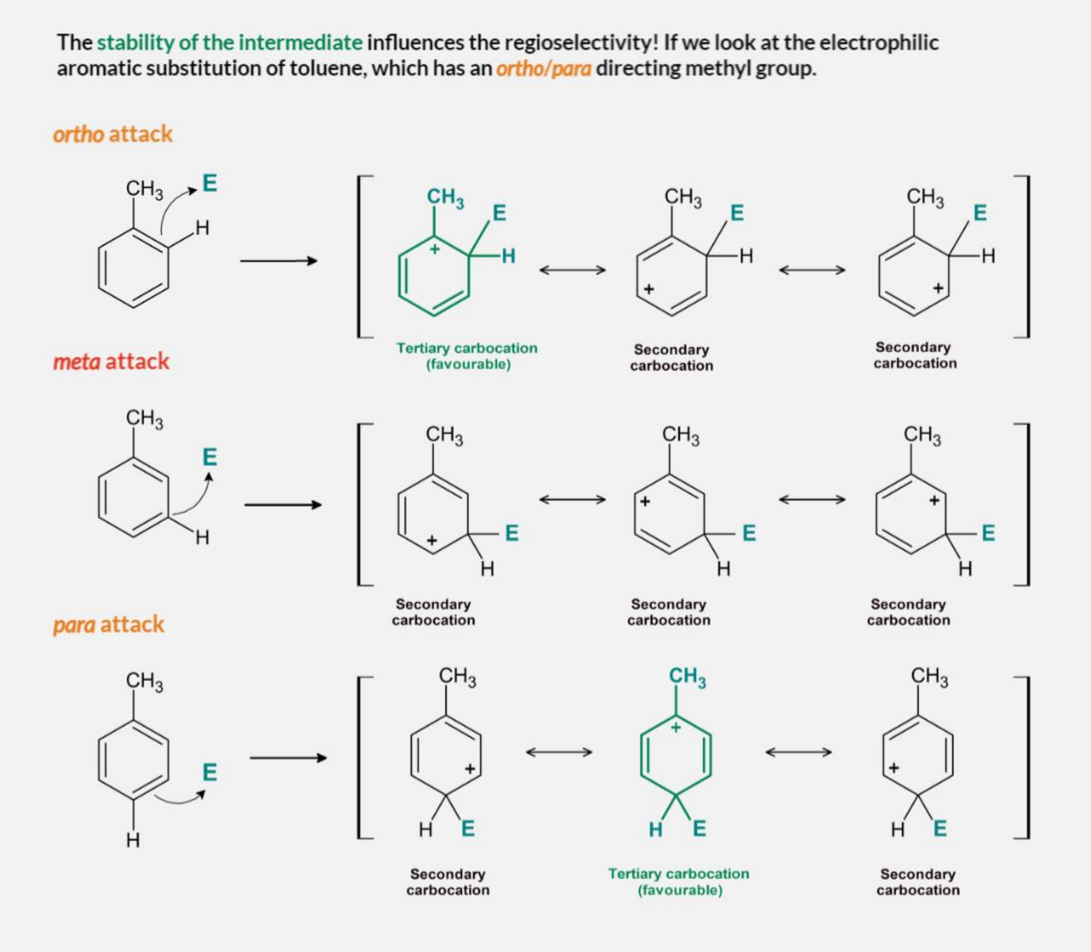



Figure 5.7. A graphical representation of intermediate stability for electrophilic aromatic substitution reactions.

Secondly, full worked examples of the electrophilic aromatic substation reaction mechanisms are provided. For this study, I focused on the Friedel-Crafts alkylation reaction as a worked example. The developed AR mechanism for this reaction is shown in figure 5.8. The creation of the AR mechanism is outlined in section 5.4.3. Addressing research question 2d, I sought to investigate the impact of AR as a visualisation aid on students' cognitive processing. I hypothesise that students using AR will report lower measures of ECL, and thus can dedicate more working memory resources to the generation of GCL. Throughout the facilitation of this educational intervention, I considered two additional points. Firstly, to provide autonomy to students, I did not impose an individual or group work setting. Regarding performance measures, no superior effects have previously been found for group work compared to individual study when utilising worked examples and instructional explanations (Kasuma and Retnowati, 2021). Yet, the individual setting imposed a lower cognitive load. In addition, previous work has found instructional explanations, and even no explanations, to be superior to self-explanations (Renkl, Atkinson, and Große, 2004). The fledgling chemist is likely to be unable to accurately diagnose their own performance deficiencies, an ability that seems to be related to an individual's knowledge of the task (Dunning, Johnson, Erlinger, and Kruger, 2003).

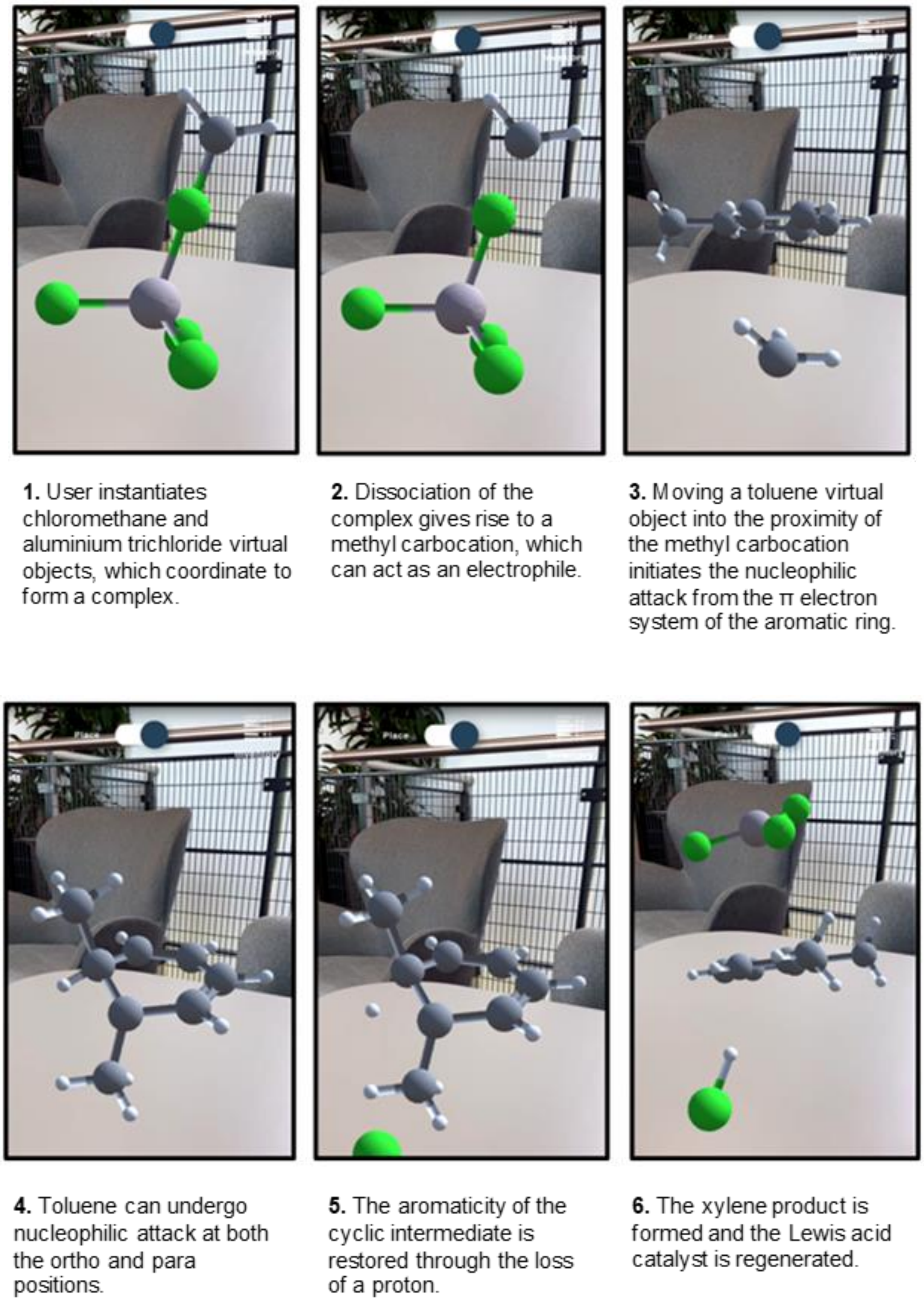



Figure 5.8. The Friedel-Crafts alkylation of toluene represented using ChemFord.

### 5.4.3 Developing immersive reaction mechanisms

One of the primary challenges of conducting this study was the development of an interactive AR electrophilic aromatic substitution reaction mechanism which could be used to support the learning of participants in the AR experimental group. Throughout this section, an overview of the development of the AR reaction mechanism is presented, which explores the virtual objects created and the underlying code driving the functionality. Within the Unity Editor, which served as the environment where the AR experience was developed, classes of code were written using Visual Studio (2022) IDE 2019. The programming language specification of VS is C#, and the C# files created have the extension *.cs*. To create the electrophilic aromatic substitution AR reaction mechanism, the molecules presented in table 5.2 were created using Blender v.2.9 (Foundation, 2022), an open-source computer graphics software toolset. For each of the molecules listed, details regarding the C# components attached to those virtual objects are also provided.

Table 5.2. Details of the virtual molecules creating alongside the C# components added.

| **Molecule (virtual object)** | **C# components** |
|---|---|
| Aluminium trichloride | GenerateHCl.cs<br>FinishFriedelCrafts.cs<br>CreateFixedJoint.cs<br>ChargedParticle.cs<br>IonManager.cs |
| Toluene | GenerateIntermediates.cs<br>IonManager.cs |
| Chloromethane | FormAlCl4.cs<br>MovingProton.cs<br>IonManager.cs |
| Xylene | - |
| Non-aromatic intermediate | CreateCCBond.cs<br>MovingProton.cs |
| Hydrogen chloride | - |

Figure 5.9 outlines the first step of the Friedel Crafts alkylation of toluene; the generation of a carbocation electrophile (shown in blue) resulting from the reaction of aluminium trichloride and chloromethane.

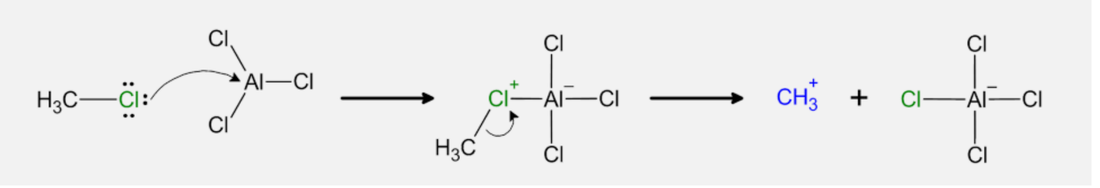


Figure 5.9. The generation of a carbocation resulting from the reaction of chloromethane and aluminium trichloride.

To replicate this first step within ChemFord, chloromethane and aluminium trichloride were constructed as virtual objects, with components added allowing them to be rescaled, moved, and rotated when manipulated by a user. However, this does not provide the functionality required to complete the reaction step shown in figure 5.9. To achieve this, three C# components were constructed and added the aluminium trichloride virtual object:

i. **CreateFixedJoint.cs**. This component is responsible for moving the chloromethane molecule into the correct position when approaching the aluminium trichloride molecule. In addition, this component is responsible for the bond breaking (C−Cl) and bond forming (Al−Cl) operations.

ii. **ChargedParticle.cs**. This component is responsible for assigning the correct positive or negative charge to a species. The code for the class is shown below.

```
using System.Collections;
using System.Collections.Generic;
using UnityEngine;

public class ChargedParticle : MonoBehaviour
{
    // Set the charge value (+/-) for the species.
    [SerializeField]
    public float charge;
}
```

iii. **IonManager.cs**. This component is responsible for managing all charged species in the AR environment and applying the correct electrostatic force. The code for this class is shown below.

```
public class IonManager : MonoBehaviour
{
    // How often this code executes (every 0.01 seconds)
    private float cycleInterval = 0.01f;
    private List<ChargedParticle> chargedParticles;
    private List<MovingProton> movingChargedParticles;

    void Start()
    {
        // Find all of the charged species in the AR environment
        chargedParticles = new
List<ChargedParticle>(FindObjectsOfType<ChargedParticle>());
        movingChargedParticles = new
List<MovingProton>(FindObjectsOfType<MovingProton>());

        foreach(MovingProton mp in movingChargedParticles)
        {
            StartCoroutine(Cycle(mp));
        }
    }

    // Check how many charged species exist in the AR environment.
    public IEnumerator Cycle(MovingProton mcp)
    {
        bool isFirst = true;
        while (true)
        {
            if (isFirst){
                isFirst = false;
                yield return new WaitForSeconds(Random.Range(0,cycleInterval));
            }
            // Apply an electrostatic force to any charged species present.
            ApplyElectrostaticForce(mcp);
                yield return new WaitForSeconds(cycleInterval);
        }
    }

    // The code for calculating the electrostatic force to apply.
    private void ApplyElectrostaticForce(MovingProton mp)
    {
        Vector3 newForce = Vector3.zero;
        foreach(ChargedParticle cp in chargedParticles)
        {
            if (mp == cp)
                continue;
            // Calculate the distance between any two charged species.
            float distance = Vector3.Distance(mp.transform.position,
cp.gameObject.transform.position);
            // MP and CP are two charged species,
            // Mathf.Pow(distance,2) means distance squared.
            float force = 1000 * mp.charge * cp.charge/Mathf.Pow(distance,2);

            Vector3 direction = mp.transform.position - cp.transform.position;
            direction.Normalize();

            // Update the electrostatic force every 0.01 seconds
            newForce += force * direction * cycleInterval;
            // If the value is not a number, set the force value to 0.
            if (float.IsNaN(newForce.x))
                newForce = Vector3.zero;
            mp.rb.AddForce(newForce);
        }
```

In addition to the IonManager.cs component, the chloromethane virtual object also contained 2 additional components:

i. **MovingProton.cs**. This component is attached to the $CH_3^+$ electrophile. When an aromatic molecule such as toluene is in close proximity, this component draws the electrophile to the *ortho* or *para* position of that species for substitution. The code for the class is shown below.

```
using System.Collections;
using System.Collections.Generic;
using UnityEngine;

public class MovingProton : ChargedParticle
{
    // Apply a mass to the virtual molecules.
    public float mass = 1;
    public Rigidbody rb;

    // If the object doesn't have a collider, create a collider.
    void Start()
    {
        if (!gameObject.GetComponent<Rigidbody>())
        {
            rb = gameObject.AddComponent<Rigidbody>();
        }
        else
        {
            rb = gameObject.GetComponent<Rigidbody>();
        }
        rb.mass = mass;
        // Remove the effects of gravity.
        rb.useGravity = false;
    }
    // If the species is a hydrogen ion, make charge +1.
    public void IsHydrogen()
    {
        gameObject.GetComponent<MovingProton>().charge = 1;
    }
}
```

ii. **FormAlCl4.cs.** This component is responsible for changing the shape of the aluminium trichloride molecule from trigonal planar to tetrahedral once it gains a chlorine substituent and becomes aluminium tetrachloride. The code for the class is shown below.

```
public class FormAlCl4 : MonoBehaviour
{
    // Refers to the virtual molecule objects
    private Transform _AlCl3;
    public GameObject _chloromethane;
    public GameObject _chloromethaneCollider;

    // The control for the shape change animation
    private Animator _anim;

    // If the two molecules come into contact, run this code.
    private void OnTriggerEnter(Collider other)
    {
        _AlCl3 = other.transform.parent;
        if (other.gameObject.CompareTag("AlCl3"))
        {
            if (_AlCl3.GetComponent<Animator>() == null)
            {
            }
            else
            {
                // Change from trigonal planar to tetrahedral.
                _anim = _AlCl3.GetComponent<Animator>();
                _anim.SetBool("TriToTetra", true);
                _anim.SetBool("TetraToTri", false);
            }
        }
        gameObject.transform.SetParent(_chloromethaneCollider.transform);
        // Destroy the chloromethane collider.
        _chloromethaneCollider.name = "ToBeDestroyed";
        gameObject.GetComponent<SphereCollider>().isTrigger = false;
        float _sphereRadius = gameObject.GetComponent<SphereCollider>().radius;
        gameObject.GetComponent<SphereCollider>().radius = _sphereRadius / 1.5f;
        _chloromethane.transform.parent = gameObject.transform;
        gameObject.transform.SetParent(null);
        other.transform.eulerAngles = new Vector3(0f, 0f, 0f);
        gameObject.transform.eulerAngles = new Vector3(-1.908f, -167.14f, 109.68f);
        gameObject.GetComponent<MovingProton>().charge = 0.005f;
        transform.localScale = new Vector3(200, 200, 200);
        other.transform.parent.localScale = new Vector3(5, 5, 5);
        // Generate an Ion Manager class for each new charged species created.
        foreach (GameObject gameObj in GameObject.FindObjectsOfType<GameObject>())
        {
            if (gameObj.name.Contains("Selection Manager"))
            {
                if (!gameObj.GetComponent<IonManager>())
                {
                    gameObj.AddComponent<IonManager>();
                }
            }
        }
    }
}
```

To trigger these C# components, colliders were added to the aluminium trichloride and chloromethane virtual objects. In AR, a collider handles a collision between virtual objects. They are invisible, and do not need to be the same shape as the object's mesh (Unity3d, 2022). For the aluminium trichloride and chloromethane virtual objects, primitive colliders (such as box and sphere colliders) were used as they are the least processor-intensive (figure 5.10). The scripting system can detect when collisions occur within the AR environment and initiate resultant actions. For this step of the reaction mechanism, the C# components initiated are the three components attached to aluminium trichloride and the two components attached to chloromethane. This is illustrated by steps 1 and 2 in figure 5.8.

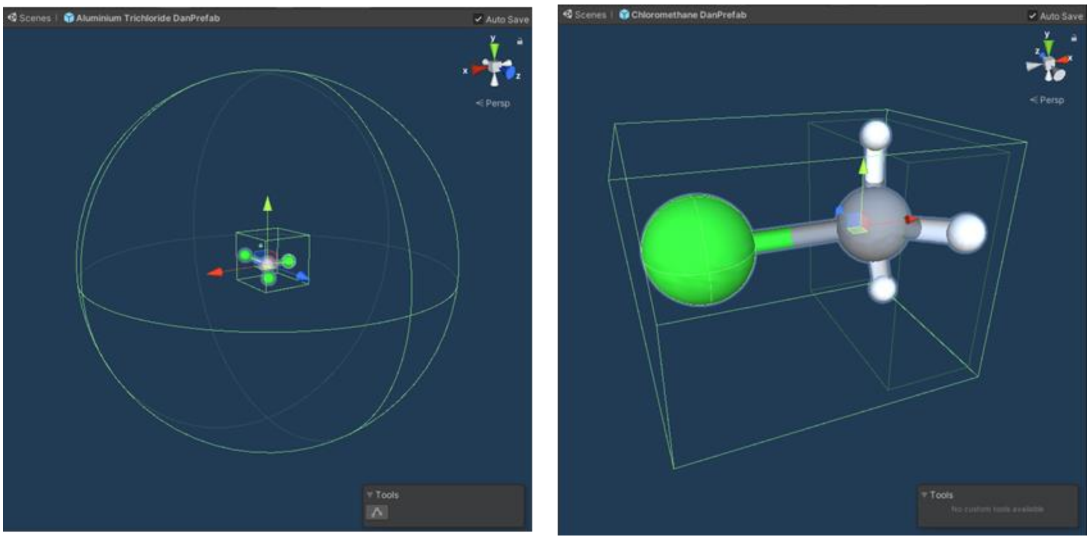


Figure 5.10. The sphere collider surrounding aluminium trichloride (left) and the box collider surrounding chloromethane (right). The interaction of these colliders initiates the attached C# components.

The next step of the reaction mechanism is the generation of a resonance stabilised non-aromatic intermediate formed from the nucleophilic attack of toluene π electrons (figure 5.11). As methyl groups are *ortho*/*para* directing, substitution can occur at two positions. As such, it was important that the AR environment replicated this behaviour. Additionally, AR affords the ability to view different resonance structures of the formed intermediate. As such, this functionality was also developed within ChemFord for this AR reaction mechanism.

This step of the AR reaction mechanism is achieved by applying a sphere collider to each of the double bonds on the toluene virtual object (figure 5.12). When a double-bond sphere collider detects a collision with the methyl carbocation, the C# component GenerateIntermediate.cs applies an electrostatic force which results in the formation of a new C−C bond using component CreateCCBond.cs. If the sphere collider at the *ortho* position detects a collision, the *ortho* intermediate is formed, else the para intermediate is generated. This is illustrated by steps 3 and 4 in figure

5.8. Resonance structures can be toggled through a button on the ChemFord user interface (UI).

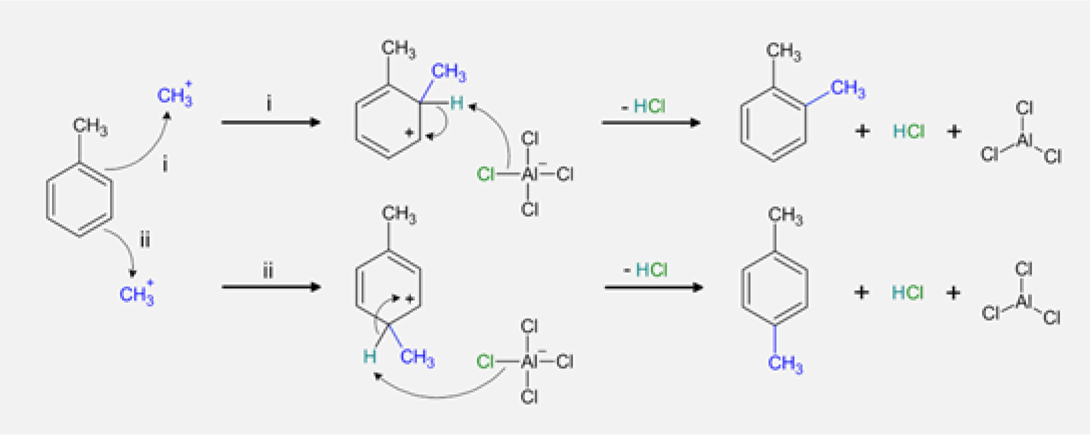



Figure 5.11. *Ortho*/*para*-substitution of a methyl carbocation on toluene, followed by the loss of a proton, which restores aromaticity.

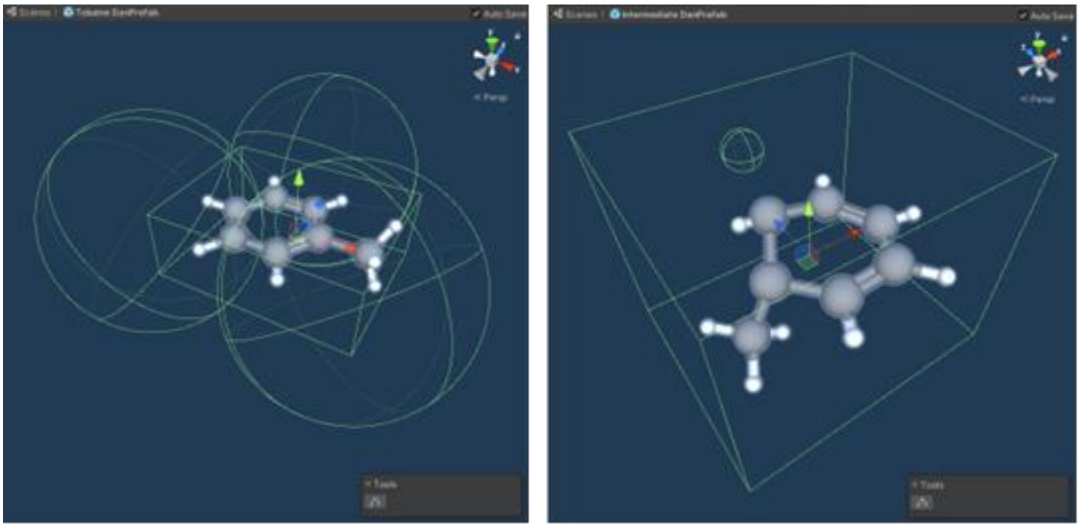


Figure 5.12. The double bond sphere colliders of toluene (left) and the ortho intermediate (right). The sphere collider on the ortho intermediate signifies the substitution position of the methyl carbocation.

Lastly, the loss of a proton restores the aromaticity, generating the corresponding xylene product. Depending on the position of substitution, this will be ortho- or para-xylene. In addition, aluminium trichloride is regenerated and hydrogen chloride is produced as a side product. All three of these operations are controlled by a single component called FinishFriedelCrafts.cs. The code for this class is shown below and is illustrated in steps 5 and 6 of figure 5.8.

```
public class FinishFriedelCrafts : MonoBehaviour
{
    // A list of the GameObjects involved
    public GameObject _HClMolecule, _AlCl4Bond4, _chlorineAnchor, _AlCl4Cl4;
    private Transform _AlCl3;
    private Animator _anim;
    public GameObject _chlorinePosition, _collider;

    private void OnTriggerEnter(Collider other)
    {
        _AlCl3 = gameObject.transform.parent;
        if (other.gameObject.CompareTag("AlkylIntHydrogen"))
        {
            // Change shape from tetrahedral to trigonal planar.
            _AlCl4Cl4.GetComponent<ChargedParticle>().charge = 0;
            _anim = _AlCl3.GetComponent<Animator>();
            _anim.SetBool("TriToTetra", false);
            _anim.SetBool("TetraToTri", true);
            _HClMolecule.SetActive(true);
            StartCoroutine(RevertToAlCl3());

            // Destroy an IonManager and MovingProton components.
            Destroy(gameObject.transform.parent.GetComponent<IonManager>());
            Destroy(other.GetComponent<MovingProton>());
            Destroy(other.gameObject);
            _chlorineAnchor.SetActive(true);
            _chlorineAnchor.AddComponent<ChargedParticle>().charge = -1;
            string _changeName =
transform.parent.GetComponent<CategoryAssignment>()._objectName = "Aluminium
trichloride";
            foreach (Transform child in _AlCl3)
            {
                if (child.name == "Chlorine 4")
                {
                    child.GetComponent<MeshRenderer>().enabled = false;
                }
            }
            // Generate the HCl molecule.
            Transform _HCl = Instantiate(_HClMolecule.transform,
gameObject.transform.parent);
            _HClMolecule.SetActive(false);
            _HCl.SetParent(null);
        }
    }
    // Revert AlCl4 back to AlCl3.
    IEnumerator RevertToAlCl3()
    {
        yield return new WaitForSeconds(2.5f);
        _AlCl4Bond4.SetActive(false);
        _chlorinePosition.SetActive(true);
        _collider.SetActive(true);
        foreach (Transform child in _AlCl3)
        {
            if (child.name == "Chlorine 4")
            {
                child.GetComponent<SphereCollider>().isTrigger = false;
                child.GetComponent<MeshRenderer>().enabled = true;
                child.gameObject.SetActive(false);
            }
        }
    }
}
```

## 5.5 Results and Discussion

Descriptive statistics concerning the measured variables of cognitive load, conceptual understanding and CAM are summarised in tables 5.3 and 5.4. These results can provide information regarding the relative effectiveness of the pedagogical intervention. Following data collection, the Shapiro-Wilk test was used to check my data for normality. Although other methods for normality testing exist, Shapiro-Wilk has more power to detect nonnormality in smaller sample sizes (Mishra et al., 2019). In addition, Bartlett's test was used to confirm that the equality of variances was true.

Table 5.3. Relative means and standard deviations for $S_E$Ar conceptual knowledge scores and cognitive load measures.

| **Variable** | Control group<br>**Mean (SD)** | AR group<br>**Mean (SD)** |
|---|---|---|
| **$S_E$Ar Test Instrument**<br>***0 (low) to 90 (high)*** | n = 22 | n = 12 |
| Pre-test | 43.33 (14.14) | 40.00 (14.14) |
| Post-test | 50.00 (17.32) | 55.56 (24.06) |
| **CLS responses**<br>***(11-point scale)*** | n = 23 | n = 18 |
| ICL | 5.91 (1.51) | 6.36 (1.65) |
| ECL | 3.41 (2.22) | 3.58 (1.56) |
| GCL | 6.44 (1.48) | 6.26 (1.72) |

34 students completed the $S_E$Ar test instrument at both the pre- and post-test stages. Data pertaining to conceptual knowledge was found to be nonnormally distributed for both the pre- and post-test stages. Consequently, intergroup comparisons of the scores on the $S_E$Ar test instrument from each experimental group were conducted using the Mann-Whitney U test. No significant differences were observed in the pre-test scores, $p$ = 0.579, or in the post-test scores, $p$ = 0.514. The calculated Cohen's $d$ value for the post-test $S_E$Ar test instrument scores was 0.14, suggesting negligible differences between the two experimental groups. However, the Wilcoxon signed-rank test (the non-parametric variant of the paired-

samples *t*-test) showed significant intragroup improvement for the AR group concerning $S_E$Ar test instrument scores, $p = 0.029$. This improvement was not observed in the control group, $p = 0.204$. Thus, for this cohort of students, the introduction of AR technology, over and above the implementation of worked examples, resulted in significant intragroup improvement regarding performance on the $S_E$Ar test instrument.

Table 5.4. Median and interquartile range for current achievement motivation measures.

| **QCM responses** ***(7-point Likert scale)*** | Control group **Median (Interquartile range)** | AR group **Median (Interquartile range)** |
|---|---|---|
| | n = 23 | n = 18 |
| Anxiety | 3.20 (1.60) | 3.50 (1.80) |
| Challenge | 4.25 (1.25) | 4.50 (1.25) |
| Interest | 4.60 (1.40) | 4.50 (1.80) |
| Probability of success | 4.50 (1.25) | 4.50 (1.00) |

## 5.5.1 $S_E$Ar Reliability Analysis

Furthermore, I also sought to establish reliability on the SEAr test instrument, to provide information regarding the quality of each item. Similar to the ITMC test instrument, discussed in chapter 3, the concepts and analytical procedures of CTT (DeVellis, 2006) and IRT (Embretson and Reise, 2000) were employed to determine item and scale difficulty and discrimination values. These values are shown in table 5.5. Again, the extreme group method was used to calculate discrimination within the groups, partitioned by the top and bottom 27% (Preacher, 2015). This is done to avoid the issues of weakening data associated with procedures such as the median split (Aiken, West, and Reno, 1991).

In the context of educational testing, a difficult item is one that more respondents answer incorrectly. The difficulty values calculated range from 0–1, where a higher value indicates an easier item. The most effective items have mid-ranges of difficulty. However, in practice, a difficulty of 0.5 on every test item for every cohort is not realistic to target. Therefore, difficulty values within a range of 0.3–0.9 are acceptable. Items more strongly correlated with other items, and thus the true score, are fundamentally better items. Such items are said to have greater discrimination. CTT analysis indicates, for this cohort, that items 3 and 4 are the easier items on the instrument, with item 5 being the hardest item. Consequently items 3 and 4 display the lowest level of discrimination between learners of lower and higher ability. The other items on the instrument display better levels of discrimination.

Table 5.5. Item difficulty and discrimination values for each of the 9 items on the $S_E$Ar test instrument.

| | CTT* | | IRT | |
|---|---|---|---|---|
| **Instrument Item** | **Difficulty** | **Discrimination** | **Difficulty** | **Discrimination** |
| 1 | 0.59 | 0.26 | −15.81 | 0.023 |
| 2 | 0.74 | 0.27 | −10.22 | 0.100 |
| 3 | 0.85 | 0.06 | −43.76 | 0.040 |
| 4 | 0.88 | 0.21 | −17.01 | 0.119 |
| 5 | 0.29 | 0.77 | 0.649 | 26.927 |
| 6 | 0.53 | 0.79 | −0.112 | 3.796 |
| 7 | 0.56 | 0.71 | −0.220 | 1.963 |
| 8 | 0.44 | 0.70 | 0.172 | 2.036 |
| 9 | 0.56 | 0.56 | −0.300 | 1.013 |

* CTT is dependent on the sample population.

IRT analysis of the ITMC data was performed using dichotomous 1PL, 2PL, and 3PL models. To assess the absolute fit of each model, two measures were examined. Firstly, a generalisation of Orlando and Thissen’s (2003) S-χ2 item-fit statistic was inspected. The item-fit statistic assesses the degree of similarity between model-predicted and empirical response frequencies by item response category. A statistically significant value indicates that the model does not fit a given item. The S-χ2 fit statistic for each item (table 5.6) indicates a satisfactory fit in 8 of the 9 items for the 1PL, 2PL, and 3PL models.

Table 5.6. Item-fit statistics for 1PL, 2PL, and 3PL models. Statistically significant values are highlighted in blue.

| ITMC item number | 1PL | | 2PL | | 3PL | |
|---|---|---|---|---|---|---|
| | ***S-χ2*** | ***P*** | ***S-χ2*** | ***p*** | ***S-χ2*** | ***p*** |
| 1 | 8.05 | 0.090 | 1.40 | 0.706 | 4.42 | 0.110 |
| 2 | 6.18 | 0.186 | 3.38 | 0.337 | 3.10 | 0.212 |
| 3 | 19.60 | 0.001 | 10.39 | 0.016 | 10.44 | 0.005 |
| 4 | 0.50 | 0.974 | 0.75 | 0.862 | 0.64 | 0.727 |
| 5 | 6.49 | 0.165 | 3.79 | 0.286 | 4.88 | 0.087 |
| 6 | 3.50 | 0.479 | 2.58 | 0.461 | 2.46 | 0.293 |
| 7 | 3.43 | 0.488 | 4.32 | 0.229 | 3.95 | 0.139 |
| 8 | 4.40 | 0.356 | 6.68 | 0.083 | 5.18 | 0.075 |
| 9 | 0.84 | 0.933 | 0.81 | 0.847 | 1.38 | 0.503 |

The fit of the 2PL and 3PL models to the data were compared using the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) fit statistics (table 5.7). For the two statistics, a lower value indicates a better model fit to the data (Acquah, 2010). The addition of the pseudo-guessing parameter (3PL model) did not improve the model fit. Thus, the 2PL model was used to interpret item parameters.

Table 5.7. The model level fit comparison for 2PL and 3PL models for this study.

| Model | Log-Likelihood | AIC | BIC |
|---|---|---|---|
| 2PL | −161.85 | 359.70 | 387.18 |
| 3PL | −159.63 | 373.26 | 414.47 |

Figure 5.13 displays the item-characteristic curves (ICC) generated from my 2PL model. ICCs are the fundamental unit in IRT and can be understood as the probability of answering a dichotomous item correctly, for individuals with a given ability (Goldhammer, Martens, and Lüdtke, 2017). Items that are easy to correctly answer are shifted to the left of the scale, whereas items that are difficult to answer correctly are shifted to the right. Generally, the ICC have an ogive curve, beginning on the left with low probabilities of answering an item correctly for lower values of student ability, rising to represent increasing probabilities of answering the item correctly as student abilities increases. Items on the instrument displayed good

discrimination, constituting reasonable evidence that each item's score is positively related to the overall proficiency represented by performance on the $S_EAr$ test instrument. Items 1−4 were considered the easiest items, generally at the lower estimate of individuals' ability. The inflection points of these items lie at an ability lower than −4. Thus, these items are easier and discriminate less. As such, I have omitted them from the item characteristic curves shown in figure 5.13.

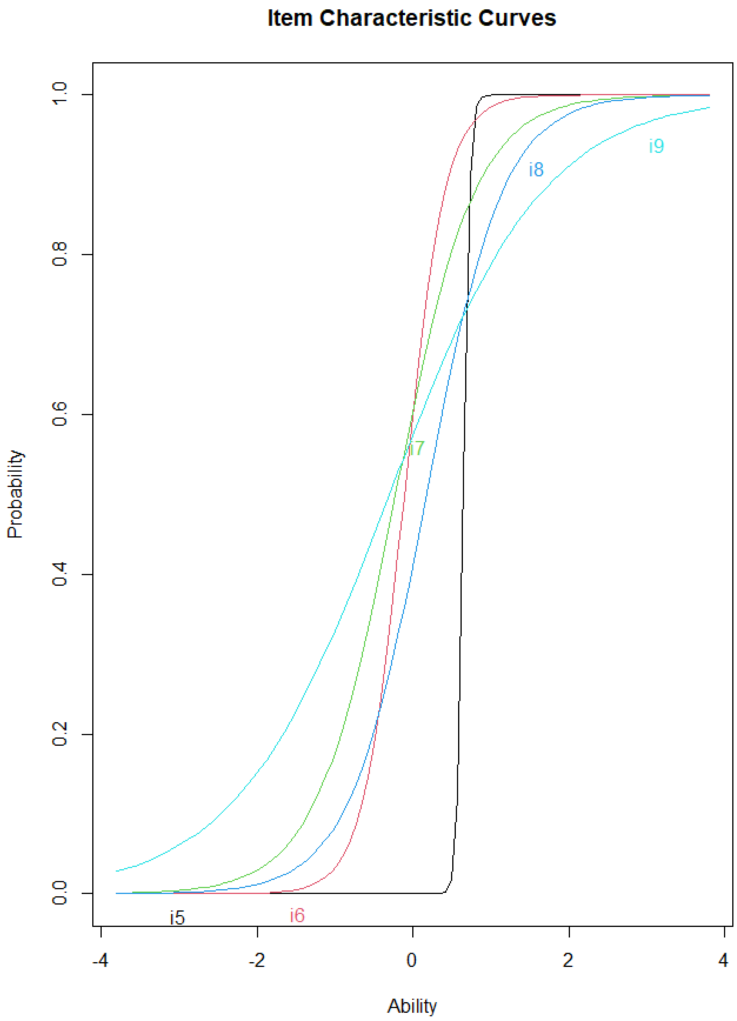



Figure 5.13. The item characteristic curves for items on the ITMC, generated using a 2PL model, excluding items 1−4.

The range of values suggests that the $S_EAr$ instrument contains items of varying difficulty, appropriate for differentiating between students studying the topic of electrophilic aromatic substitution. Items 5–9 demonstrate difficulty values around the mean of the population distribution for ability. Discrimination values constitute reasonable evidence that each item's score is positively related to the overall proficiency represented by performance on this instrument. Internal consistency of the instrument was determined through calculation of Cronbach's alpha. The alpha value for the $S_EAr$ instrument was 0.62, acceptable for an assessment used for low-stakes purposes (Cortina, 1993). The removal of any items would not lead to any substantial gains in Cronbach's alpha. This suggests each item coheres with the rest of the test.

I employed Differential Item Functioning (DIF; Karami, 2012) to see if students of equal ability, but from different groups, have unequal probability to respond correctly to the items on the $S_EAr$ test instrument (table 5.8). This is because DIF items can lead to biased measurement of ability. The DIF is stated to be uniform or non-uniform depending on whether the discrepancy in item performance between subgroups is consistent or non-consistent respectively. If an item is identified as having DIF, this is due to a source of variance not related to the structure measured by the test (Messick, 1994). DIF studies have an important role in assessing the validity of test scores (Finch and French, 2015) as the presence of DIF in the test items may reduce the validity of the test.

For this study, I employed the Raju Signed Area Method. The detection thresholds used were –1.96 and 1.96, with a significance level of 0.05. Item 6 was detected as DIF.

Table 5.8. DIF analysis conducted on items of the $S_EAr$ test instrument.

| Item # | Statistic | *p* value | DIF detected? |
|---|---|---|---|
| 1 | –0.090 | 0.9285 | NO DIF |
| 2 | 0.413 | 0.6800 | NO DIF |
| 3 | –0.059 | 0.9532 | NO DIF |
| 4 | 0.213 | 0.8312 | NO DIF |
| 5 | –0.014 | 0.9887 | NO DIF |
| 6 | –2.271 | 0.0232 | DIF |
| 7 | –0.882 | 0.3777 | NO DIF |
| 8 | –0.913 | 0.3614 | NO DIF |
| 9 | 0.096 | 0.9234 | NO DIF |

### 5.5.2 CLS and QCM data

A total of 41 students completed the CLS instrument. Prior to analysis of the CLS data, the existence of normality and equality of variances was confirmed using the Shapiro-Wilk test and Bartlett's test. Intergroup comparisons for each component of cognitive load were carried out using independent samples t-tests. No significant differences were detected for ICL, t(39) = 0.903, $p = 0.372$, ECL, t(39) = 0.292, $p = 0.772$, or GCL, t(39) = 0.361, $p = 0.720$. I hypothesised that the introduction of AR would assist students' mental visualisation, thus reducing the reported measures of ECL, whilst achieving similar or improved scores on the $S_E$Ar test instrument in the post-test stage. As ECL decreases, more working memory resources are available to deal with ICL, maximising the generation of GCL. To account for participants' prior knowledge, a one-way ANOVA was employed, introducing the pre-test scores obtained on the $S_E$Ar test instrument as a covariate. For ICL, $F(1,29) = 0.112$, $p = 0.741$, and ECL, $F(1,29) = 0.989$, $p = 0.329$, tests of between-subject effects showed no significant differences in cognitive load. GCL approaches significance with Bonferroni post hoc comparisons showing higher levels of GCL in the AR group, $p = 0.098$.

Cronbach's alpha calculations were conducted as a measure of internal consistency of the CLS instrument, with values of 0.92 for the ICL and GCL sub-scales and 0.89 for the ECL scale. This demonstrates very good internal consistency. Furthermore, measures of ECL negatively correlated with measures of GCL, $r(41) = -0.600$. As extraneous load, imposed by suboptimal instructional design increases, effective learning decreases. This relationship, calculated using Pearson's correlation, was significant at $p = 0.01$. Internal consistency measures for the four sub-scales of the QCM were also calculated and indicated acceptable Cronbach's alpha values of 0.59 (challenge), 0.77 (interest), 0.75 (probability of success), and 0.72 (anxiety). Using Spearman's correlation, I observed that the dimensions of interest, challenge, and probability of success all positively correlated with GCL. This was significant at $p = 0.01$ (table 5.9). Interest has been previously shown to be an important predictor of test performance (Freund, Kuhn, and Holling, 2011).

Table 5.9. Relationship between GCL and QCM measures.

| | QCM Measure | | | |
|---|---|---|---|---|
| | Challenge | Interest | Probability of success | Anxiety |
| **GCL** | 0.517* | 0.548* | 0.336* | 0.008 |

* Correlation is significant at the 0.01 level (2-tailed)

In addition, measures of ECL were found to be negatively correlated with interest, $r(41) = -0.482$, and challenge, $r(41) = -0.492$. This was again significant at $p$ = 0.01. The levels of extraneous cognitive processing increased as students' interest and perceived difficulty of the activity decreased. Lastly, measures of probability of success were negatively correlated with ICL, $r(41) = -0.297$. As element interactivity increased, student's perception of the probability of surmounting the task decreased. This was close to reaching significance at $p$ = 0.05. For measures of anxiety, interest, and probability of success, a one-way ANCOVA showed that pre-test scores were not related to these measures. However, for the challenge measure, Bonferroni post hoc pairwise comparisons approach significance, $p$ = 0.082, with higher values for the AR group. Thus, students in the AR group may have perceived the learning task as easier when utilising AR technology and were therefore more motivated towards completing the challenging tasks.

## 5.5.3 Cognitive, Affective and Performance Measures

Normalised change ($c$) calculations were conducted as a measure of the learning gain of students between the pre- and post-test stages. The higher the normalised change, the greater the learning gain. For this study, the ranges defined by Hake (1998) for normalised gain are adopted: low ($c < 0.3$), medium ($0.3 \leq c < 0.7$); and high ($0.7 \leq c$).

Firstly, regarding the two different modes of representation, $c = 0.12$ for the control group and $c = 0.22$ for the AR group. In addition, the extreme group method was used to differentiate between students of lower and higher prior relevant chemistry experience. Groups were partitioned by the top and bottom 27% (Preacher, 2015). For students with lower relevant chemistry experience, $c = 0.30$, whereas for students displaying higher relevant chemistry experience, $c = 0.10$. Students with lower prior chemistry experience demonstrated greater learning gains when interacting with the worked examples activity. Furthermore, the reported measures of ECL (for students of lower and higher relevant prior chemistry experience) were compared to their calculated normalised change. For participants exhibiting lower prior relevant conceptual knowledge, the mean value of ECL = 2.94, whereas for participants with higher prior relevant conceptual knowledge, the mean value of ECL = 4.40. This difference was shown to be approaching statistical significance, $p$ = 0.095. Thus, an expertise reversal effect; described as the reversal of the effectiveness of instructional techniques on learners with differing levels of prior knowledge (Kalyuga, 2009), is present. Previous findings support the view that example-problem pairs may be more effective for learners with lower prior knowledge (Reisslein, Atkinson, Seeling and Reisslein, 2006; van Gog, Kester and Paas, 2011). When comparing the modes of representation of the worked examples (2D vs AR), intergroup comparisons of normalised change between experimental groups shows no significant differences, $p$ = 0.585. However, a medium effect size was calculated ($d$ = 0.25).

Regarding motivational measures, no significant differences were found for the four sub-scales of the QCM instrument between participants demonstrating lower and higher mean scores on the $S_EAr$ test instrument: interest ($p$ = 0.366), probability of success ($p$ = 0.968), anxiety ($p$ = 0.844), and challenge ($p$ = 0.424). In addition, no significant differences were found between groups when introducing pre-test $S_EAr$ test scores as a covariate. The association between measures of ECL, and the four sub-scales of the QCM instrument, for students of lower and higher prior relevant chemistry experience, is shown in table 5.10. This was calculated using Spearman's correlation. The two groups were again partitioned using the extreme group method.

Table 5.10. Spearman's correlation values ($r_s$) calculated between ECL and QCM measures, for participants of lower and higher prior relevant chemistry experience.

| | | **QCM Measure** | | | |
|---|---|---|---|---|---|
| | Group | Challenge | Interest | Probability of success | Anxiety |
| **ECL** | Low | −0.312 | −0.564 | −0.294 | 0.332 |
| | High | −0.815* | −0.824* | −0.262 | −0.091 |

* Correlation is significant at the 0.01 level (2-tailed)

For participants with lower prior relevant chemistry experience, anxiety was found to be positively correlated with probability of success, $p = 0.05$. In addition, measures of probability of success were found to be strongly positively correlated with measures of challenge. This was significant at the $p = 0.01$ level.

For participants with lower prior conceptual knowledge, anxiety was found to be positively correlated with probability of success, $p = 0.05$. In other words, anxiety may contribute as a motivator to drive students to surmount the learning activity, a relationship reported in previous works (Strack, Lopes, Esteves and Fernandez-Berrocal, 2017). Furthermore, measures of probability of success were found to be strongly positively correlated with measures of challenge, $p = 0.01$. Regarding cognitive load measures, ECL was found to be negatively correlated with measures of student interest; a relationship that was approaching significance, $p = 0.056$. For participants of higher prior relevant chemistry experience, ECL was strongly negatively correlated with both challenge and interest. As elements of the learning material become redundant, students' perceived difficulty of the activity and their interest in completing the learning activity decreases. Lastly, measures of challenge were strongly positively correlated with interest, $p = 0.01$, and moderately positively correlated with probability of success. This relationship was approaching significance, $p = 0.068$. Instructional design, tailored at the appropriate difficulty for the learner will spark interest, hopefully supporting successful completion of the learning experience.

### 5.5.4 Analysis of Qualitative Data

I recruited 10 students in total, from both experimental groups, to participate in semi-structured interviews. The interview schedule covered two topic areas:

i. Students' perception and satisfaction in response to engaging with my worked example learning activity.
ii. A discussion based on topics of electrophilic aromatic substitution.

Qualitative analysis of the participant interviews was completed through latent thematic analysis using the approach of Braun and Clarke (2006). Data was recorded, and transcribed verbatim, prior to being subjected to analysis for commonly occurring themes. The initial broad themes were constructed based on frequency and similarity of responses. Redundancy was eliminated and closely related major themes were merged. For this study, I focus on two predominant themes found in student discussions: designing effective worked examples; and students' understanding of $S_E$Ar. Study interviewees (SI), for purposes of pseudonymisation, are again represented by a number.

I sought to ensure reliability in my analysis using negotiated agreement. The extent of agreement between coders was measured using Krippendorff's alpha. Two researchers independently coded the full set of interview transcripts and then negotiated how they applied the codes. Differences were discussed and where there was a consistent disagreement, a common approach was agreed (the negotiated codebook employed can be found under Appendix L). Krippendorff's alpha is a commonly used chance-corrected reliability measure that avoids many of the limitations described for Cohen's kappa, such as its suitability to smaller samples sizes (Krippendorff, 2018). Krippendorff's alpha has ranges between −1.00 and 1.00, with positive values indicating agreement beyond chance. Values above 0.66 are acceptable for tentative conclusions (Krippendorff, 2018). The Krippendorff's alpha calculated for this set of coded interview transcripts was 0.82.

The first theme pertains to **designing effective worked examples**. In their accounts, participants highlighted their views of, and experiences with, my worked examples. To avoid confounding the potential benefits afforded by my AR tool, and to minimise sources of ECL, design principles of the CTML were employed. My quantitative data suggests that interest is strongly negatively correlated with measures of ECL. This was reflected in participants' responses, in terms of positive student satisfaction:

*"I really like the booklet, the collection of examples. The step-by-step layout in which it was given. I really liked it, I wanted to take it with me after that session"* (SI 2); and in terms of the worked examples supporting the learning process:

*"It's really good to fall back to for reference if I ever forget any of the steps or any of the core ideas"* (SI 1).

*"The fact that it's step by step, that it's broken up into steps... So, first the mechanism, then the substituents etc. The fact that it's structured in a way that you can follow easily..."* (SI 3).

Regarding the design of visual elements within my worked examples, evidence of CTML principles were noted in students' accounts: *"Breaking it down into smaller chunks is a lot easier"* (segmenting principle; SI 1); "...the description at the start is just really concise. It's down to the point" (coherence principle; SI 4); *"It made it very visually easy to read. It wasn't just, you know, blocks of text..."* (multimedia principle;

SI 6); and *"...that's really clear. Right next to it is the activating and deactivating groups. I find this table really handy"* (spatial contiguity principle; SI 5). The integration of my AR tool was also very positively perceived. Participants commented that the visualisation affordance of the technology supported their learning: *"The actual model of it, though, I thought was really good. I thought it was brilliant to be fair, importing the chemicals, and seeing the 3D view in front of you"* (SI 4).

To both enhance learning and improve comprehension, we used colour to direct attention and associate information. Students' responses indicated that this assisted retrieval practice when answering the faded problems. This is a finding in line with previous studies (Dzulkifli and Mustafar, 2013).

*"...the colour coding helped understand it. It was well laid out"* (SI 7).

In contrast, a minority of participants noted that the use of colours may be a possible source of distraction, and hence ECL, potentially diminishing the generation GCL. *"There are quite a few different font colours. I think some people find that quite distracting..."* (SI 1). However, the psychology of colour, and its impact on the visual elements embedded within the learning activity, is outside the scope of this study.

Regarding element interactivity, participants implied that mentally processing a worked example, containing all steps of an electrophilic aromatic substitution reaction, may overwhelm their working memory. *"I think maybe if it had been broken up a bit more, so maybe a bit of information and then a question about the information. Then more information followed by another question"* (SI 3). Interlinking smaller worked examples for each step, paired with faded problems, that subsequently lead to a larger faded problem that encapsulates multiple steps may be a more effective approach for tasks considered to be of higher element interactivity. Lastly, the inclusion of an introduction to electrophilic aromatic substitution theory, provided by the facilitator prior to participants attempting the worked examples, was noted as an important step to this pedagogical approach. *"I would rather be taught a chunk of material and then given this to reinforce it, you know, to really drive home, the mechanisms and stuff like that"* (SI 10).

The second theme identified throughout my thematic analysis is **students' understanding of $S_E$Ar**. Participants' understanding of the concepts underlying $S_E$Ar, in response to completing the worked examples, and faded problems, were explored. Students could identify examples of both activating and deactivating groups: *"I could quite easily go for methyl [substituents] being activating and nitro [substituents] being deactivating"* (SI 1); *"Ester groups are deactivating"* (SI 8). In addition, students demonstrated sound understanding of what constitutes activating and deactivating groups:

*"Activating groups are able to donate electron density into the pi orbitals above and below the aromatic ring. That will be things like amine groups... Deactivating is when they pull electron density away from the ring structure. So, that will be cyanide groups and nitro groups"* (SI 5).

I expanded my discussions to analyse how students convey the effects activating and deactivating groups have on the $S_EAr$ reaction. Regarding the rate of reaction, participants could explain that activating groups *"...are going to increase the rate"* (SI 8); *"It increases it [the rate of reactivity]"* (SI 10). In addition, interlinking the influence of activating and deactivating groups on substitution position, students recalled that: *"So, activating groups tend to be ortho/para, and deactivating groups tend to be meta"* (SI 3), but also provided evidence of deeper understanding:

*"So, I know, if it's electron donating, it's more like to be ortho/para. And if it's electron withdrawing, it's more likely to be meta"* (SI 9). In terms of regioselectivity, students exclaimed that substitution position will be a result of *"...the groups attached to it [the ring] and where the charge ends up."* (SI 1).

Moreover, discussions were extended, on the influence of attached functional groups on substitution position, to include disubstituted aromatic molecules. Throughout students' accounts, three common responses were apparent:

- The more activating group will control the position of substitution: *"It would be the more activating group. It would be the methyl group"* (SI 8)
- Steric effects will primarily dictate the position of substitution: *"The nitro group? It's bulkier. Right?"* (SI 2)
- The group that is more activating or deactivating will control the position of substitution: *"I feel like it will be the nitro group because the nitro group is more strongly deactivating than the methyl group is activating."* (SI 7).

Next, my discussions shifted to focus on the $S_EAr$ reaction mechanism, in which I focused on three distinct areas:

i. Changes in aromaticity
ii. The role of the Lewis acid.

iii. The rate determining step.

To start, I asked students to comment on whether any changes in aromaticity occur throughout the $S_EAr$ reaction, and, if so, at what point(s). A majority of participants could accurately explain that a loss of aromaticity is initially observed on the ring system as a result of the bonding of the electrophile: *“You'd lose the aromaticity of the ring.”* (SI 6), and that the aromaticity is regenerated through deprotonation: *“When the intermediates form, technically it's lost, but I mean, it regains it”* (SI 4).

Regarding the role of the Lewis acid and the rate determining step, a majority of students conveyed reasonable understanding. Participants could correctly identify the role of the Lewis acid catalyst:

*“...it’s deprotonating”* (SI 1).

*“...you end up with a positive charge and AlCl4 which attacks the hydrogen to take it away”* (SI 3).
In addition, a majority of students could successfully identify the rate determining step: *“It will be the formation of the intermediate”* (SI 5); *“It’s the original breaking of the aromaticity to form the tetrahedral carbon”* (SI 9).

Following on, the discussion transitioned from $S_EAr$ concepts to specific examples of $S_EAr$ reactions. Students could identify both the Friedel-Crafts alkylation and acylation, in addition to examples such as:

*“I remember the nitration. So, sulfonation and nitration...”* (SI 4).

*“...chlorination and bromination...”* (SI 5).

*“I think the Vilsmeier-Haak mechanism was mentioned”* (SI 7).

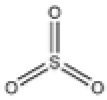
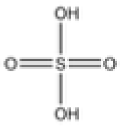
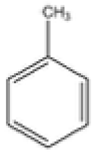
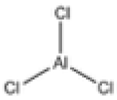
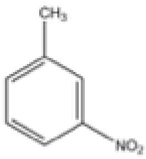
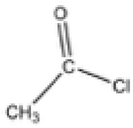
Remaining on the topic of Friedel-Crafts alkylation, I captured discussion points regarding carbocation rearrangement and unwanted subsequent reactivity. Most students recognised that carbocation rearrangement occurred within alkylation reactions, but only a minority of students were able to disclose the reason why: *“...it rearranges to be... it would prefer to be secondary or tertiary, it’s more stable”* (SI 5); and that *“with acylation, it will always be primary, as [it’s] an acyl chloride”* (SI 8). Further, only a small number of participants demonstrated understanding of the limitations of polyalkylation: *“So, with regards to alkylation, [methyl] groups increase the electron density and thus increases the reactivity towards electrophiles”* (SI 4). A common misconception was that this reactivity was caused by interactions with the

Lewis acid catalyst: *"It could be a problem because it could potentially react with the AlCl4?"* (SI 8).

Finally, I presented students with two visual elements containing three molecules (table 5.9). For each example, I asked the following questions:

- Which molecule will be acting as the nucleophile, the electrophile and the Lewis acid catalyst in a $S_EAr$ reaction?
- What is the name of the $S_EAr$ reaction being displayed?
- Where will the new group be substituted, with respect to the aromatic starting reagent.

Table 5. 11. Examples 1 and 2 shown to participants throughout the semi-structured interviews.

| | Molecule 1 | Molecule 2 | Molecule 3 |
|---|---|---|---|
| Interview example 1 |  |  |  |
| Interview example 2 |  |  |  |

Example 1 was answered well by most participants, who were able to distinguish that this reaction was a sulfonation, and that the major product observed would be substitution at positions ortho/para to the methyl group of toluene. In contrast, when discussing example 2, students would commonly attribute it as a chlorination reaction. From further probing, it was apparent that participants struggled to identify that an acyl chloride functional group was present. I corrected for this, and most students revaluated that the reaction was in fact a Friedel-Crafts Acylation. Most participants could correctly assign the substitution position of the incoming electrophile for example 2: *"It's going to be at positions one and three"* (SI 10); *"...NO2 is deactivating, and the methyl group is activating... So, it's going to be ortho and para to the CH3 group"* (SI 5).

## 5.6 Limitations

Some limitations of this study must be acknowledged. Firstly, a major limitation is the relatively small sample size that the data analysis was based upon. The sample size was the result of modest enrolment compounded by participant disengagement between the pre- and post-test stages. For instrument reliability analysis using approaches such as CTT, larger sample sizes are preferable where possible. In addition, I must acknowledge the possibility of self-selection bias from participants (Heckman, 1990). Students who volunteer for interviews may be different from the rest of the population regarding their communication ability or reasoning levels. Lastly, the absence of a delayed post-test, for conceptual understanding, prevents the evaluation of long-term retention.

## 5.7 Chapter Conclusions

Instructional guidance, such as that provided by worked examples, helps the novice learner deal with complex information, that may be difficulty to process in limited capacity working memory. This study illustrates how worked examples, adopting the affordances of AR technology can support the learning of electrophilic aromatic substitution. Referring to research question 1, regarding measures of cognitive load and achievement motivation, no significant differences were observed between experimental groups. This was unaffected when introducing prior relevant chemistry experience as a covariate. QCM measures of challenge, interest, and probability of success were found to correlate positively with reported GCL. Reported ECL negatively correlated with reported GCL, in addition to measures of challenge and interest. Measures of challenge and interest demonstrated a stronger negative correlation with ECL for students displaying higher prior relevant chemistry experience.

Regarding research question 2, no significant differences were observed between groups for conceptual understanding, demonstrated by the mean scores achieved on my $S_EAr$ instrument, at both the pre- and post-test stages. Yet, significant intragroup improvement and greater normalised change values were observed for the AR group. No significant intragroup improvement was found in the control group for conceptual understanding. Initial reliability analysis for the $S_EAr$ instrument was conducted using CTT and IRT. Items 1–4 are generally at the lower estimate of individuals' ability, whereas items 5–9 demonstrate difficulty values around the mean of the population distribution for ability.

In an attempt to answer research question 3, it was found that participants displaying higher prior conceptual knowledge also reported higher measures of ECL, alongside lower normalised change values. As learner expertise increases, a shift to a heavier emphasis on problem solving may be beneficial. For learners with lower relevant chemistry experience, challenge was strongly correlated with probability of success. Commenting on research question 4, student feedback and subsequent thematic analysis showed that the developed worked examples, alongside implementation of my AR tool, were positively perceived by students. Commenting on research question 4, the qualitative data suggests how CTML design principles may have supported learning, as well as how participants conveyed their understanding of $S_EAr$ concepts following my intervention.

# 6

# AR meets Peer Instruction

Difficult chemistry concepts are challenging to teach. For the expert chemist, the mastery of a difficult concept results in a perspective shift (Cousin, 2006), where previously eluding propositions seem almost self-evident. This cognitive bias, famously coined “The Curse of Expertise” (Hinds, 1999), can result in fluency misattribution. Consequently, an expert intuitively assumes that fundamental pieces of information, obvious to them, are also obvious to the fledgling chemist. To the experienced educator this is clearly not the case. Previous works, designed around active learning, have been developed in an attempt to promote conceptual understanding through the interactive engagement of students (Hake, 1998). One approach, first reported by Mazur (1997), is the student-centred pedagogy of Peer Instruction (PI). Eric Mazur developed PI in the 1990s at Harvard University, initially for use in large, introductory physics classrooms. Yet, the method is now used within a variety of disciplines at a range of institutional levels. PI engages students during class through structured, frequent questioning, facilitated by classroom response systems, to resolve misunderstandings. The central feature of PI is the ConcepTest (CT), a conceptual, multiple-choice question (MCQ), designed to help resolve common student difficulties around the subject content.

Within this chapter, the fourth educational intervention, a coordination chemistry PI session, supported by augmented reality (AR) is presented. Within this session, students were provided two opportunities to answer each of my developed questions – once after a round of individual reflection, and then again after a round of AR-supported peer instruction. The second round provides students with the opportunity to “switch” their original response to a different answer. Typically, the proportion of correct responses increases after peer discussion. For the six questions posed, I analysed students’ discussions, in addition to their interactions with the ChemFord AR tool. Furthermore, students’ self-efficacy, and how this, in addition to factors such as CT difficulty, influence response switching was examined.

An introduction to the format of PI is presented in **section 6.1,** followed by a discussion of self-efficacy in **section 6.2**. Details regarding the development of the six CTs posed throughout the PI session are outlined in **section 6.3.** In addition, information pertaining to the experimental design employed throughout the study, including details of the test instruments employed is outlined in **section 6.4**. Next, the design process of the AR experiences developed to support students' discussions around coordination chemistry is presented in **section 6.5**. In **section 6.6**, the results of my qualitative analysis are presented. Evidence of resource activation, in terms of knowledge elements and control structures, constituted the main themes of my thematic analysis. As such, an introduction to the three resource types, as defined by Tuminaro and Redish (2007), in addition to the concept of Epistemic Games are also provided within this section. The quantitative results of the study are presented in **section 6.7.** Within this section, data pertaining to CT difficulty, in addition to a reliability analysis conducted using Item Response Theory (IRT) is reported. The reliability analysis was conducted to provide users with information regarding the quality of each question. Descriptive statistics outlining the extent to which students switched their answers between the first and second round of voting, are presented in **section 6.8**. Following, an examination of students' self-efficacy measures is discussed in **section 6.9**. The limitations identified within this study are discussed in **section 6.10**, with concluding remarks presented in **section 6.11**.

## 6.1 Peer Instruction

Within a PI session, time is organised by a sequence of questioning, interactive discussion, and explanation (Schell and Mazur, 2015). Following a brief introduction, the focus of the session rapidly shifts from the facilitator to the student through posing a CT. After a short period of independent thinking, students are asked to commit to an answer. This is frequently referred to as the first round of voting. Previous reported qualitative evidence has highlighted the pedagogical importance of this step in the generation of high-quality peer discussion. Nicol and Boyle (2003) report that 82% of students indicate a preference for answering the question individually before engaging in peer discussion, with comments suggesting that the individual response time forced them to think about and identify an answer to the question (Nicol and Boyle, 2003). Subsequently, students stated that this increased their engagement during peer discussion, and that discussion after a round of individual thinking resulted in deeper thinking. In contrast, omitting round 1 and starting with peer discussion has been found to frequently result in greater levels of passivity, and thus, lower levels of critical thinking. Nielsen et al. (2016) report comparable results, in that the first round of individual voting was necessary for students to form opinions without the risk of being influenced by other students. Next, the first set of answers are tallied. If too few students responded correctly during the first round of voting, the concept, in which the CT was based, is revisited. In contrast, if a large majority of students responded correctly, a brief explanation of the concept is provided, and the facilitator moves onto the next topic. Lastly, if an appropriate proportion of the cohort answers correctly, students are encouraged to engage in peer discussion, and then revote on the CT (figure 6.1). This is referred to as the second round of voting.

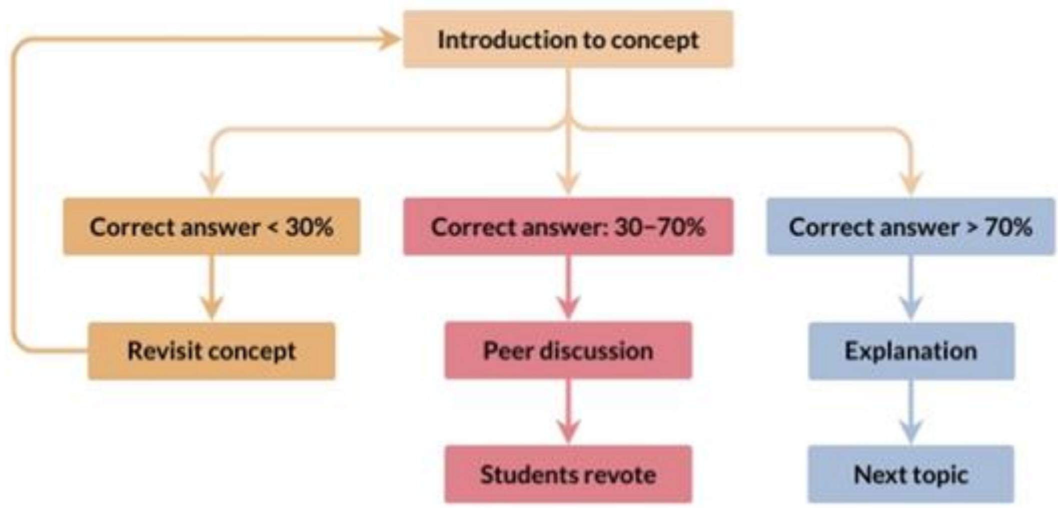



Figure 6.1. The PI implementation procedure, adapted from Mazur (1997).

PI is considered to be a low threshold pedagogy as it can be implemented into an educational environment with little effort or technological support. The element of peer discussion is arguably the most recognizable feature of the PI model and works to maximise both the amount of time that students think about key concepts, in addition to the time students spend engaging in self-monitoring of their understanding of the discipline. As students explain their understanding of a CT, often an epiphany occurs, which takes them further than their individual thinking processes. This is similar to the zone of proximal development, which refers to the difference between what an individual can learn alone, and what they can achieve with guidance from others, such as an instructor or peers (Bransford, Brown, and

Cocking, 2000; Vygotsky, 1978). In round 1, students work individually, before participating in deeper learning through peer discussion in round 2. As such, one important aspect of PI is to clarify that the observed improvement is more than students simply following those who are correct. Smith et al. (2009) found in their statistical analysis that students who answered a question incorrectly, but the following isomorphic question correctly, did not belong to a discussion group with a student who knew the correct answer. Hence, these students were presumably able to arrive at the correct answer through peer discussion. Further researchers have replicated these results in the disciplines of computing (Porter et al., 2011) and general chemistry (Bruck and Towns, 2009).

Further hypothesises to explain the effectiveness of PI have also been suggested, such as the greater amount of time allowed for individual reflection and information processing. Lasry et al. (2016) investigated whether other metacognitive processes, such as reflection or time spent on a task, explained the learning gains associated with PI. All participants engaged in the first round of voting, and then engaged with one of three tasks during round 2:

i. Peer discussion.
ii. Silent reflection on answers.
iii. Distraction using a cartoon.

The learning gains of students, following round 2, were highest when students engaged in peer discussion. Works such as those discussed previously constitute only a small sample of more than 20 years of research on the use of interactive teaching methods such as PI to improve student learning. The body of research on PI, primarily from physics education researchers indicates that PI significantly improves student learning outcomes, such as conceptual understanding and problem-solving ability. Improvements in student learning and engagement when PI is used has been reported in other science (Golde, McCreary, and Koeske, 2006; Knight and Wood, 2005; Smith et al., 2009), technology, engineering (Nicol and Boyle, 2003), and maths (STEM; Miller, Santana-Vega, and Terrell, 2006) disciplines.

As such, implementation of the process outlined in figure 6.1 has provided compelling evidence that PI is associated with substantial improvements in students' ability to solve conceptual and quantitative problems (Mazur, 1997; Vickrey et al., 2015). Empirically, normalised learning gains twice as large as those associated with traditional lectures have been observed when implementing PI effectively (Crouch and Mazur, 2001). Students' academic performance is a fundamental indicator when implementing educational interventions, and one popular measure is normalised learning gain, first introduced by Hake (1998). As a measure, normalised gain allows valid comparisons of learning between students with diverse levels of prior conceptual knowledge.

Yet, academic performance is well established as complex, with numerous variables contributing simultaneously. Such dimensions may be cognitive, and are well studied (Kuncel and Hezlett, 2010), whilst others reside in the affective domain. Within this study, I focus on the affective domain factor of perceived self-efficacy, an individual's belief that one can successfully complete a task (Bandura, 1977). Self-efficacy has been previously recognised as a strong predictor of performance in science (Andrew, 1998; Pietsch, Walker, and Chapman, 2003). Students with higher self-efficacy are reported to experience fewer negative emotions in the face of difficulty, compared to students with lower reported self-efficacy (Bartimote-Aufflick et al., 2016). Self-efficacy has been shown to influence cognition, motivation, and affective processes, which in turn, can influence future self-efficacy beliefs (reciprocal determinism; Bandura, 1977). Thus, self-efficacy can impact several factors relevant to learning in a PI environment, such as perseverance and self-regulated learning (Trujillo and Tanner, 2014).

## 6.2 Self-Efficacy

Self-efficacy was first developed as an integral part of social cognitive theory (SCT), an agentic perspective to human development, adaptation, and change. As there are different social cognitive theoretical perspectives, the focus for this study is limited to the social cognitive theory proposed by Bandura (1986, 1997, 2001). SCT posits that learning occurs in a social context with a dynamic and reciprocal interaction of the person, environment, and behaviour (Bandura, 1986). Within this triadic reciprocality, each set of influences on human functioning affects the others and is in turn affected by them. The pivotal feature of SCT is the importance of social influence, and its emphasis on external and internal social reinforcement. SCT considers the unique way in which individuals acquire and maintain behaviour, while also considering the social environment in which individuals perform the behaviour. Such self-regulative capabilities to direct an individual's thoughts and actions, to attain goals, is critically important for developing a sense of agency (Schunk, and Usher, 2012).

The goal of SCT is to explain how people regulate their behaviour through control and reinforcement to achieve goal-directed behaviour that can be maintained over time. People are not simply acted upon by external forces but choose to place themselves in environments that they believe are conducive for their learning. Past experiences will influence reinforcements, expectations, and expectancies, all which shape whether, and why, a person engages in specific behaviour.

The construct of self-efficacy within SCT refers to the level of a person's confidence in his or her ability to successfully perform an action. It is a dimension of success that has been linked to student learning and perseverance (Bandura, 1986; Britner and Pajares, 2001). Social cognitive theorists emphasize that learning is most effective when peers learn from others, who are both like themselves, and display high levels of self-efficacy (Schunk and Pajares, 2005). Self-efficacy beliefs affect whether individuals think in self-enhancing or self-debilitating ways. For example, students who feel competent about performing well in mathematics (high self-efficacy—personal) are apt to engage in effective learning strategies that will benefit their learning (behavioural), as well as demonstrating greater persistence (Schunk and DiBenedetto, 2016; Schunk and Usher, 2012). Meta-analyses have been conducted on studies with diverse experimental and analytical methodologies applied across diverse spheres of functioning (Moritz, Feltz, Fahrbach and Mack, 2000; Stajkovic, Lee and Nyberg, 2009). The accumulated evidence confirms that efficacy beliefs contribute significantly to the quality of human functioning.

## 6.3 ConcepTest Development

The centrepiece of PI is the CT; a question designed to assess students' understanding of the principal concepts underlying the learning material (Lancaster, Cook, and Massingberd-Mundy, 2019; Mazur, 1977). However, what makes a CT different from a question? CTs attempt to elicit, confront, and resolve student misconceptions. Within a low-stakes environment, CTs promote higher-order thinking, allowing students to demonstrate cognitive skills that are conduits to learning. When an educator asks students to respond to CTs, they are being asked to think about what they know, and what they do not know. Yet, more importantly, they are being asked to discuss their understanding of critical concepts. This peer-to-peer interaction provides an opportunity for students to practice self-regulatory skills such as self-reflection, an integral part of learning (Lim, Ab Jalil, Ma'rof, and Saad, 2020). CTs also give students extensive retrieval practice (Halpern and Hakel, 2003), the act of generating the same information in different applications to promote long-term memory. The best CTs are those that test a student's ability to transfer their understanding to new contexts. As such, CTs may be considered equivalent to comprehension, application, or analysis questions as defined by Bloom's Taxonomy (McConnell, Steer, and Owens, 2003).

Rao and DiCarlo (2000) report that CTs can generally be classified as testing either recall, comprehension and application, or synthesis and evaluation skills. The researchers report a statistically significant difference in the proportion of students who answered the CT correctly following peer discussion for all three types of question. However, the greatest improvement was observed on questions testing synthesis and evaluation skills. This contributes to one of the most frequent questions among PI researchers:

*How do I ensure that students are engaged in discussions regarding the CT?*

Several considerations were taken throughout the development of the CTs to engage students with my educational intervention. The first was to pose CTs that were at a level of desirable difficulty for students, ideally so that a range of 30–70% correct answers after the first round of responses is observed. Further discussion regarding CT difficulty is presented in section 6.7. Regarding the implementation of AR technology into PI, a limited number of previous works are reported (Ravna, Garcia, Themeli, and Prasolova-Førland, 2022). Although VR is commonly preferred for multiuser collaboration, the role of AR for collaboration is increasing (Lukosch et al., 2015). As such, I focused on how the affordances of AR could be leveraged to promote the discussion of important topics. Throughout the development process, six CTs were generated to probe students' comprehension of coordination chemistry concepts. CTs 1−3 are discussed in this section. For CTs 4−6, please see section 6.6.3. The first CT is presented in figure 6.2. CTs were reviewed for content validity to ensure that students' attention was focused on the critical concepts, which are key to addressing specific learning goals. To satisfy these requirements, I used the following six criteria when creating each CT (Newbury, 2014):

i. **Clarity**. Students should waste no cognitive resources understanding the requirements of the question.
ii. **Context**. The question should be appropriate for the learning material.
iii. **Learning outcome**. The question should allow students to demonstrate that they grasp the concept.

iv. **Distractors**. The distractors should be plausible solutions to the question.

v. **Difficulty**. The question should not be too easy or too hard.

vi. **Stimulates thoughtful discussion**. The question should engage students and incentivise thoughtful discussion within their peer groups.

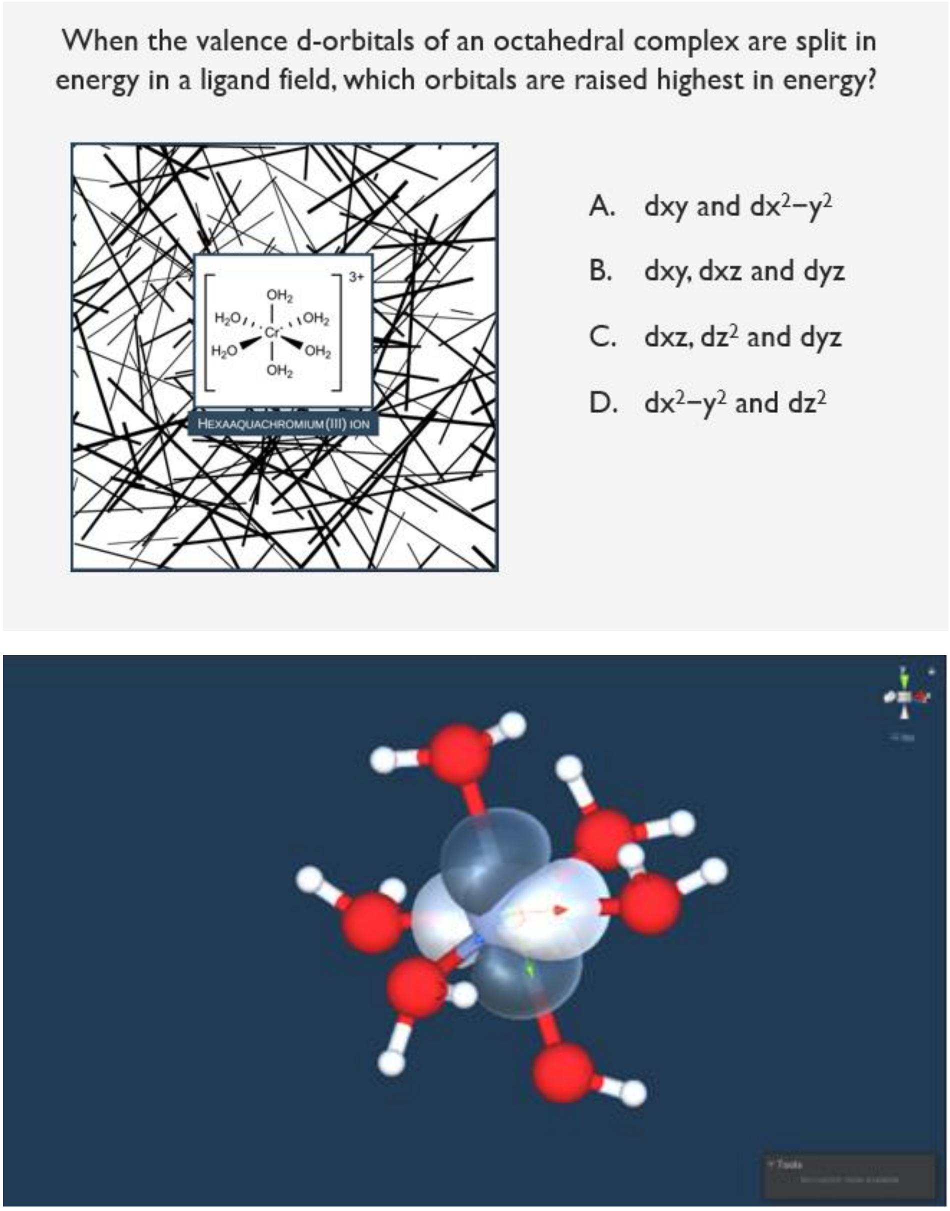



Figure 6.2. The first CT posed within my PI session (top) and the hexaaquachromium(III) ion virtual object with superimposed $dx^2{-}y^2$ orbital (bottom).

To answer the first CT correctly, there are three conceptual points which, fundamentally, students must understand:

i. Firstly, students must recognise how the axial and equatorial aqua ligands are situated around the chromium metal atom.

ii. Secondly, students must be able to comprehend the shapes and orientations of the five d orbitals of the chromium metal atom.
iii. Lastly, students must be able to comprehend the consequence of ligand and chromium d orbital interactions along the three cartesian axis (x, y, and z).

AR technology affords users the ability to instantiate interactable three-dimensional (3D) representations of the octahedral coordination sphere of the chromium complex, in addition to the 5 d orbitals of the chromium metal atom, to direct peer discussion towards these three conceptual points. The d orbitals could be easily toggled through ChemFord's user interface (UI). I envisaged that these affordances would support the generation of meaningful dialogues. Similar design aspects and considerations were implemented into my second CT, which is presented in figure 6.3.

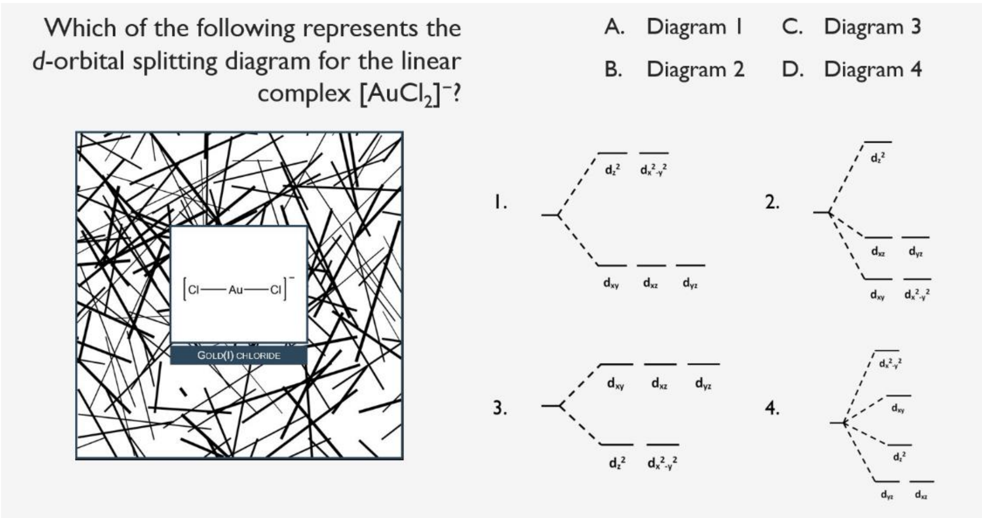



Figure 6.3. The second CT utilised within my PI session.

The second CT developed requires students to apply similar conceptual points to the first CT to arrive at the correct answer. Like the hexaaquachromium(III) virtual object, virtual d orbitals can be superimposed over the linear gold(I) chloride metal complex. As such, students are afforded the ability to inspect how the different d orbitals of the gold metal atom interact with the chlorido ligands, which lie in the z-Cartesian axis. Again, this design was intentional to promote discussions whilst also supporting students' ability to infer that the z-component d orbitals are higher in energy due to ligand-orbital interactions (specifically the $dz^2$, dxz, and dyz atomic orbitals).

Expanding on the visualisation affordance of AR technology to generate static virtual representations is the capability of AR to generate dynamic, interactable objects. As discussed in chapter 4, a trigonal bipyramidal virtual object was developed which can exhibit the Berry pseudorotation vibrational mode, an extension to the suite of VSEPR geometry virtual objects. Similarly, for my third CT (figure 6.4), a virtual object of metal complex tetraamminediaquacopper(II) was created, which was able to distort as described by the Jahn-Teller effect. The Jahn-Teller effect is a geometric distortion of a non-linear molecular system that reduces its symmetry and energy (Jahn and Teller, 1937). This distortion is typically observed among octahedral complexes where the two axial bonds can be shorter or longer than those of the equatorial bonds (Veidis, Schreiber, Gough, and Palenik, 1969). As such, elongation Jahn-Teller distortions occur when the degeneracy is broken by the stabilisation (lowering in energy) of the d orbitals with a z-component, while the orbitals without a z-component are destabilized. Such a distortion always has the effect of lowering the energy of the system to a small extent, and is thus energetically favourable (Halcrow, 2013). Within the AR environment, students could toggle the Jahn-Teller distortion of tetraamminediaquacopper(II) through ChemFord's UI to assist with CT3.

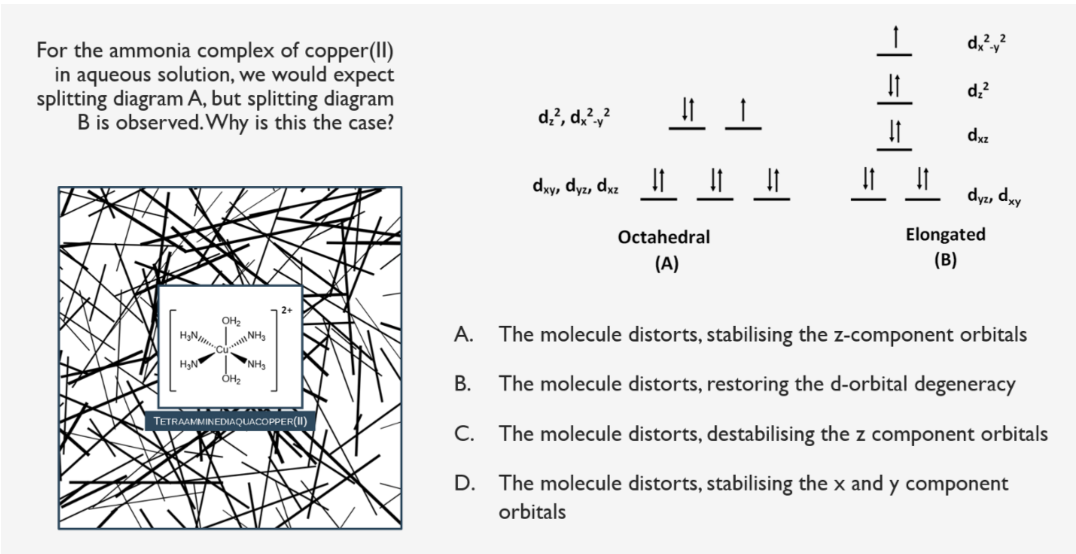



Figure 6.4. The third CT utilised within my PI session.

## 6.4 Experimental Design

The evaluation of my fourth educational intervention was conducted throughout academic year 2021/2022 as part of a FHEQ level 5 module of compulsory inorganic chemistry study at the University of East Anglia (UEA). For this study, a pre-test/post-test experimental design was employed (figure 6.5). Unlike the previous three educational interventions, which had employed comparative designs using multiple experimental groups, all participants utilised AR during the second round of each CT. The primary focus of this study was to examine if resource activation was present when students' discussions were supported by the affordances of AR technology. Student response (voting) data for my six CTs were collected through TurningPoint (now known as PointSolutions), an audience response system. In parallel, students' PI discussions, alongside their interactions with ChemFord, were captured using audio- and screen-recording software installed on a suite of iPads. This allowed the study of learning from two perspectives:

i. Probing the conceptual understanding of students through the collection of voting data.
ii. Studying the process of conceptual development during AR-supported peer discussion, through recorded conversations.

The underlying hypothesis of this work is that thinking about the qualitative dialogue, in terms of activated resources supported by AR, alongside the quantitative measures, in terms of CT difficulty and self-efficacy, can give insights into how conceptual development takes place. The research questions investigated were as follows:

**Research question 1f.** How does CT difficulty influence students' responses?

**Research question 2e.** How does self-efficacy influence students' CT responses?

**Research question 3d**. How does the integration of AR support students' PI discussions, and what types of interactions are occurring?

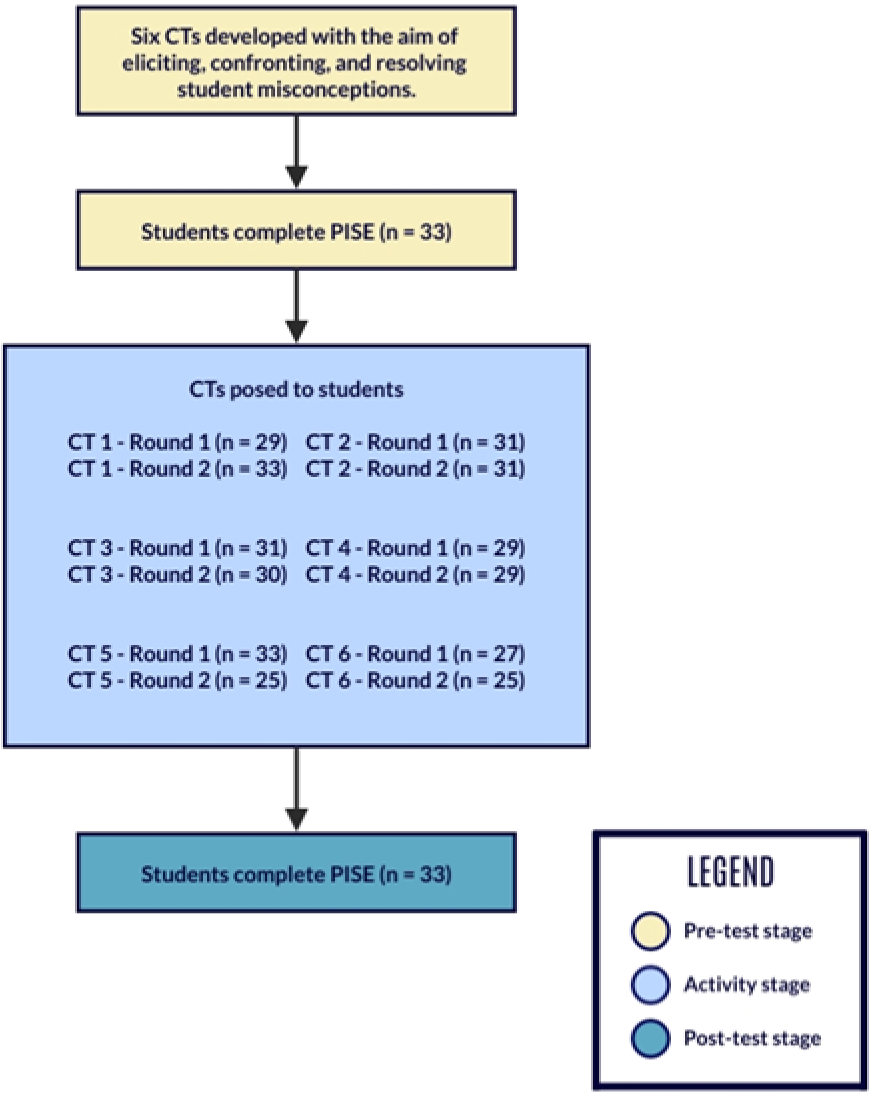



Figure 6.5. The experimental design employed, with details of participant engagement.

Regarding research question 2e, the Peer Instruction Self-Efficacy Instrument (PISE) was employed to collect students' measures of self-efficacy. The PISE is a 21-item instrument scored on a five-point Likert scale (Miller et al., 2015). The PISE was developed based on the Sources of Self-Efficacy in Science Courses survey (SOSESC; Fencl and Scheel, 2004), and Bandura (1997). For the PISE survey responses collected throughout the course of this study, I calculated Cronbach's alpha values of 0.88 for the pre-test stage, and 0.90 for the post-test stage. This demonstrates very good internal consistency. For both alpha values calculated, the removal of adapted item 12: "I get a sinking feeling when I think of trying to tackle difficult chemistry problems", resulted in a higher alpha-if-deleted value.

Previous works suggest that there is a relationship between students' performance on CTs and individual student characteristics or students' prior knowledge. As such, the true impact of PI cannot be released without controlling for student characteristics (Theobald and Freeman, 2014). A further consideration regarding the experimental design was the grouping of students. Research reports that grouping students by different responses, after the first round of voting, may lead to cognitive dissonance (Festinger, 1957). Cognitive dissonance is induced when a person holds two contradictory beliefs. As such, during peer discussion, students may aim to resolve this. In contrast, students who have the same answer, regardless of whether it is correct or not, may simply agree. For this study, students were given the freedom to form groups ranging between 2−4 students but remained within their groups throughout the duration of the intervention. This allowed control of variables such as self-efficacy and TurningPoint responses when considering the qualitative data generated through students' discussions. TurningPoint does not include a grouping tool intelligent enough to cue students to discuss their responses with peers who have different answers for each CT, without losing this capability.

The structure of my PI session is outlined in figure 6.6. Following Mazur's PI implementation procedure, students were asked to first answer independently to each CT posed. The responses were tallied, and shared with the cohort, followed by reposing of the CT. For the second round, I encouraged the use of ChemFord within students' discussions. The AR experience could be initiated through scanning an image target, which were embedded into the stem of the CT. After a period of discussion, students were asked to revote, and feedback was provided where necessary. The use of CTs and PI compares favourably to other active learning methods as rapid feedback is possible with this technique. This is especially true for instructors using electronic classroom response systems (McConnell et al., 2006; Greer and Heaney, 2004) that can be programmed to display histograms of class responses before and after peer instruction.

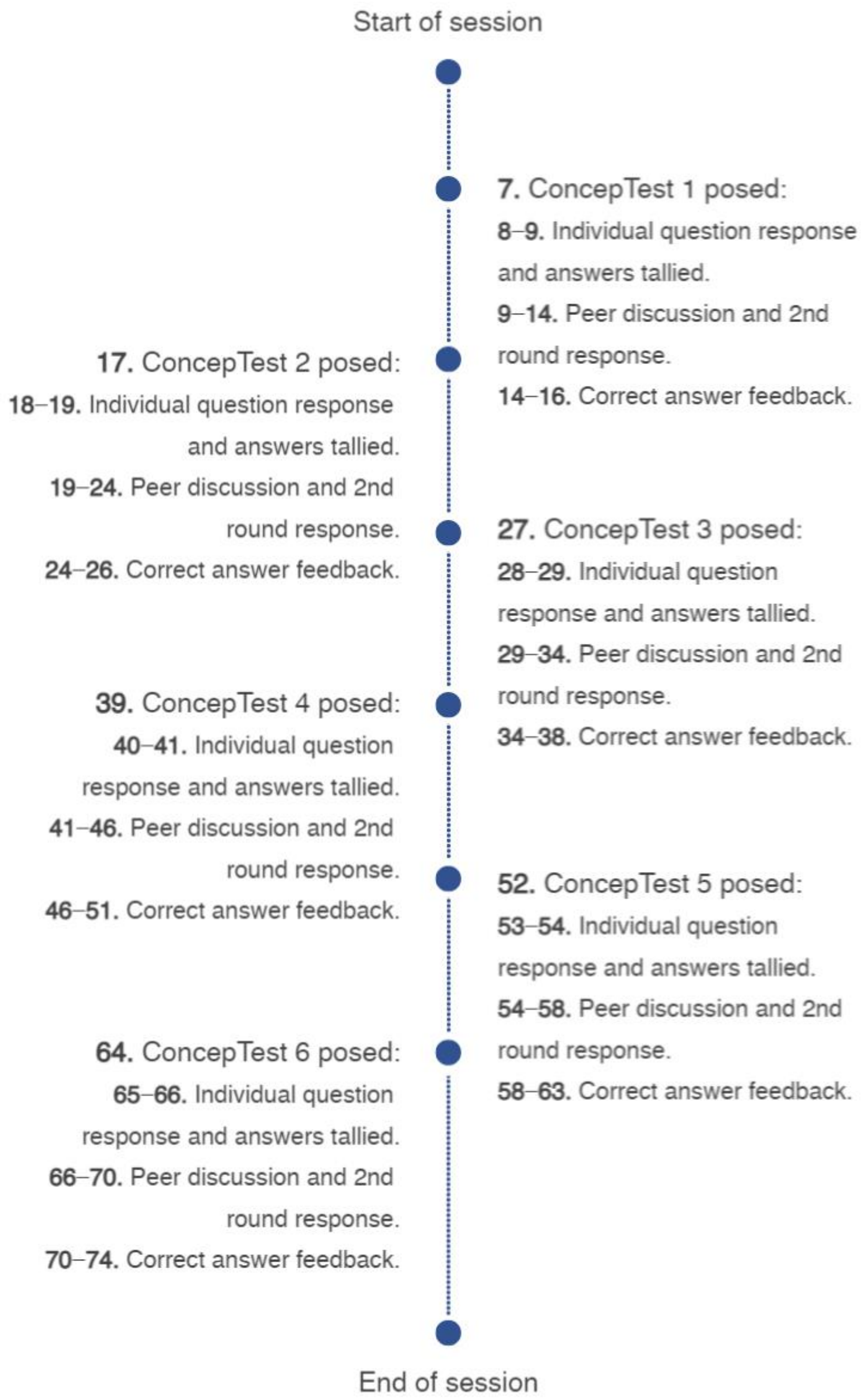



Figure 6.6. A timeline of my PI session. Numbers preceding each action indicate the session time in minutes.

## 6.5 Developing orbitals in AR

This section provides an overview of the development of different atomic and molecular orbitals which can be instantiated within my AR environment. In addition to Blender (Foundation, 2022) and the Unity Editor (2022), Gaussian 16 (2022a) and GaussView 6 (2022b) were also employed to successfully create these virtual objects. Gaussian 16 is the latest version of the Gaussian series of electronic structure programs, providing a wide-ranging suite of the most advanced modelling capabilities available. This includes predictions of energies, molecular structures, vibrational frequencies, and molecular properties of compounds and reactions. Further, GaussView can be used to visualise the atomic and molecular orbitals of a chosen system through the provision of wavefunction information. Of interest to me was not only the capability of Gaussian 16 and GaussView to generate and visualise atomic and molecular orbitals, but also the ability to export the orbital data sets into Blender as Virtual Reality Modelling Language (VRML) files. Through this, I modelled a suite of atomic orbitals that can interact to generate the corresponding molecular orbitals, depending on factors such as phase and symmetry. To achieve this, a series of chemical systems were modelled within Gaussian 16 (table 6.1) at different separation distances, measured in angstroms (Å), to capture how the atomic orbitals "morph" prior to becoming molecular orbitals.

Table 6.1. The chemical systems modelled in Gaussian 16.

| System | Bonding Interaction | Separation distance (Å) | Atomic (AO) and Molecular Orbitals (MO) generated |
|---|---|---|---|
| Diatomic hydrogen ($H_2$) | Sigma (σ) | 1.80 – 0.74 | 1s AO, s-s bonding and antibonding MOs |
| Hydrogen fluoride (HF) | | 4.00 – 0.91 | 2s and 2p AOs, s-p bonding and antibonding MOs |
| Diatomic fluorine ($F_2$) | | 4.00 – 1.43 | p-p bonding and antibonding MOs |
| Ethene ($C_2H_4$) | Pi (π) | 5.00 – 1.40 | p-p bonding and antibonding MOs |
| Dimolybdenum ($Mo_2$) | | 5.00 – 2.00 | d-p bonding and antibonding MOs |
| | Delta (δ) | 5.00 – 2.00 | 3s and 3d AOs, d-d bonding and antibonding MOs |

Within Blender, shape keys were used to achieve the morphing of AO virtual objects into their respective MOs. A shape key is used to deform an object into a new shape for animation. It is not possible to add or remove vertices in a shape key. As such, a consideration throughout development was the balance between generating data sets of sufficient density (represented by .cub files in Gaussian 16) and the representational fidelity of the morphing animations. If the data sets were too dense, computational capability requirements in the AR environment increases, whereas smaller data sets resulted in lower representational fidelity. The functionality provided by the shape keys was *baked* into each atomic orbital object prior to being exported into the Unity Editor using Autodesk Filmbox format (.fbx). Baking is the act of pre-computing functionality to improve the efficiency of other processes, as rendering from scratch is extremely time-consuming.

In Unity Editor, each atomic orbital object was wrapped in a box collider. The exceptions were the 1s, 2s, and 3s atomic orbitals, which were wrapped in sphere colliders due to their shape. In addition, each atomic orbital object was also given a rigidbody component. This puts the motion of the virtual object under the control of Unity's physics engine and is a prerequisite for collision detection. Figure 6.7 outlines the high-level logic used for creating the p-p sigma interaction within my AR environment using the Unity Editor. Users within my AR environment can instantiate atomic orbitals using marker-based or markerless AR approaches, which can then be brought into proximity of one another to construct molecular orbitals. The intersection of any two colliders within 3D space will initiate the morphing, controlled by a method called OnEnterCollision(). Details regarding the code driving this functionality is discussed further in section 6.5.2.

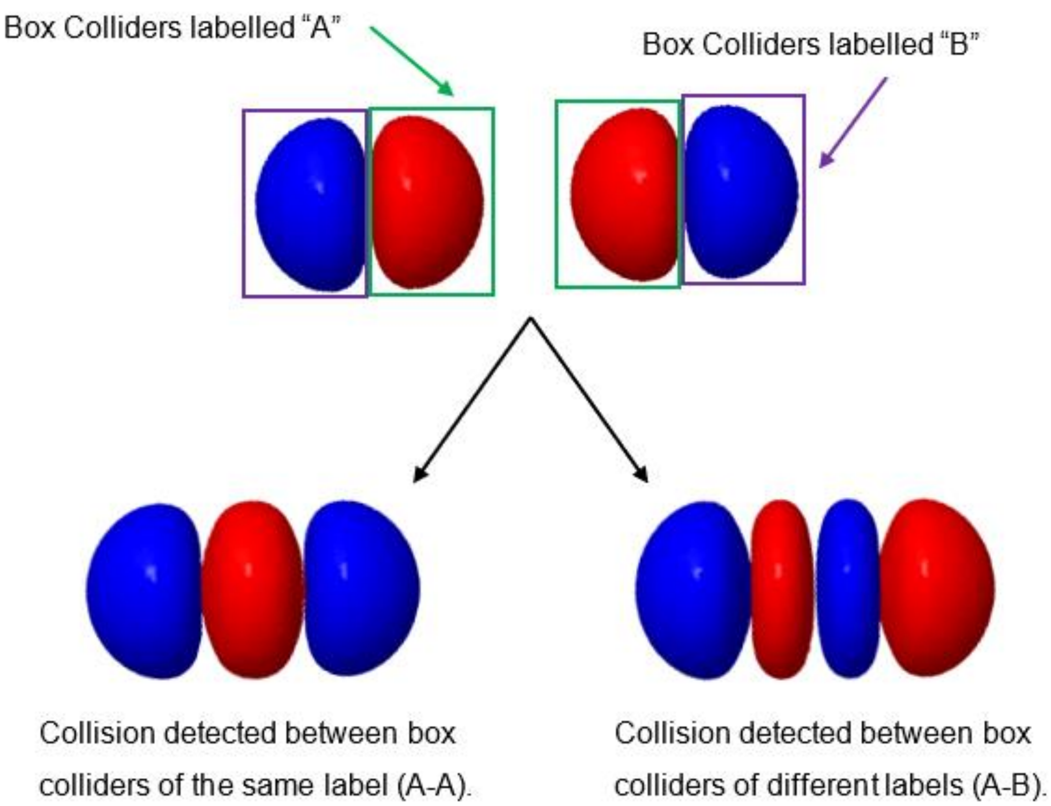



Figure 6.7. An overview of how molecular orbitals are constructed in my AR experience from the corresponding atomic orbital virtual objects.

### 6.5.1 Constructing Gaussian Z-Matrices

To model each of the molecular systems outlined in table 6.1, a series of Z-matrices were constructed. A Z-matrix is used to define the connectivity between atoms within a system. In addition, parameters such as bonding distances, angles, and dihedral angles are also defined. Each line of the Z-matrix provides the internal coordinates for each atom within the molecule/system, using the following syntax:

*Element-label, atom 1, bond-length, atom 2, bond-angle, atom 3, dihedral-angle.*

The *element-label* is a character string pertaining to either the chemical symbol for the atom or its atomic number. For systems with multiple atoms of the same type, it is customary practice to append a secondary identifying integer to the element name. As an example, consider the Z-matrix for ethene (figure 6.8):

```
1.    #n B3LYP/3-21G Opt
2.
3.    Ethene
4.
```

Figure 6.8. Lines 1−4 of the Z-matrix of ethene.

The first line of the input file specifies the basis set, a set of functions (called basis functions) that are used to represent the electron wave function in the Hartree-Fock method (HF, Fischer, 1987) or Density Functional theory (DFT, Hohenberg and Kohn, 1964). Whereas traditional HF methods attempt to find approximate solutions to the Schrödinger equation of *N* interacting electrons moving in an external, electrostatic potential, there are serious limitations of this approach (Blinder, 2019):

i. The problem is highly non-trivial, even for small values of *N*.
ii. The computational effort increases very rapidly with increasing values of *N*.

A different approach is taken in DFT, where, instead of the many-body wave function, the one-body density is used as the fundamental variable (Blinder, 2021). Gaussian 16 offers a wide variety of DFT models (Gaussian, 2022c). In HF theory, the energy has the form (equation 6.1):

$$E_{HF} = V + \langle \mathrm{hP} \rangle + 1/2\langle \mathrm{PJ(P)} \rangle - 1/2\langle \mathrm{PK(P)} \rangle \tag{6.1}$$

The terms have the following meanings:

| | |
|---|---|
| V | The nuclear repulsion energy. |
| P | The density matrix. |
| ⟨hP⟩ | The one-electron (kinetic plus potential) energy. |
| 1/2⟨PJ(P)⟩ | The classical coulomb repulsion of the electrons. |
| -1/2⟨PK(P)⟩ | The exchange energy resulting from the quantum (fermion) nature of electrons. |

Gaussian basis sets are identified by abbreviations such as N-MPG*, where *N* is the number of Gaussian primitives used for each inner-shell orbital. The hyphen indicates a split-basis set where the valence orbitals are double zeta. The double zeta basis set consists of two basis functions per atomic orbital. The *M* indicates the number of primitives that form the large zeta function, whereas *P* indicates the number that form the small zeta function.

The function with the large zeta accounts for charge near the nucleus, while the function with the smaller zeta accounts for the charge distribution at greater distances from the nucleus (Helgaker and Taylor, 1995). *G* identifies the set as being Gaussian.

The addition of an asterisk to this notation means that a single set of Gaussian 3d polarization functions is included. For example, 6-31G means each inner shell (1s orbital) Slater-type orbital (STO) is a linear combination of 6 primitives. Each valence shell STO is split into an inner and outer part (double zeta) using 3 and 1 primitive Gaussians, respectively.

In table 6.2, the basis-sets employed for purposes of modelling the molecular systems outlined in table 6.1 are presented. LANL2DZ (Los Alamos National Laboratory 2 Double-Zeta) is a widely used ECP type basis set was used to model metal atoms (Hay and Wadt, 1985).

Table 6.2. The basis-sets used for modelling different molecular systems.

| System | Basis-set |
|---|---|
| Diatomic hydrogen ($H_2$) | 3-21G |
| Hydrogen fluoride (HF) | 3-21G |
| Diatomic fluorine ($F_2$) | 3-21G |
| Ethene ($C_2H_4$) | 6-31G |
| Dimolybdenum ($Mo_2$) | LANL2DZ |

Referring to figure 6.9, the final keyword defined is *Opt*. This keyword requests that a geometry optimization be performed. The geometry will be adjusted until a stationary point on the potential surface is found. Figure 6.9 presents the remainder of the Z-matrix, where the values of 0 and 1 on line 5 denote the charge and multiplicity of the system, respectively.

```
5.    0 1
6.    C
7.    C   1 B1
8.    H   1 B2 2 A2
9.    H   1 B3 2 A3 3 D3
10.   H   2 B4 1 A4 3 D4
11.   H   2 B5 1 A5 3 D5
12.   Variables:
13.   B1        1.33579
14.   B2        1.08549
15.   A2      121.05033
16.   B3        1.08549
17.   A3      121.05033
18.   D3      180.00000
19.   B4        1.08549
20.   A4      121.05033
21.   D4        0.00000
22.   B5        1.08549
23.   A5      121.05033
24.   D5      180.00000
```

Figure 6.9. Lines 5−24 of the Z-matrix of ethene.

Line(s) 6 and 7 provide details pertaining to the two carbon atoms. Looking at line 7, the *1* denotes that the second carbon atom is bonded to the first carbon atom, by a bond of length *B1* (Å). Again, line 11, which refers to a hydrogen atom, is bonded to the second carbon atom, by a bond of length *B5*. Values *A5* and *D5* refer to the bond angle (°) and dihedral angle (°) respectively. For ethene, values of *B1* were modified between the ranges shown in table 6.1 to model the molecular orbitals of ethene, in addition to the atomic orbitals of the two separated methylene ($CH_2^+$) cations. The same approach was applied to the other four molecular systems.

## 6.5.2 Constructing dynamic orbital virtual objects

Figure 6.10 shows the 1s orbital. In addition to the shape of the orbital, toggleable axis were also associated with these objects. Furthermore, the sphere collider can be seen, represented by the green concentric lines around the sphere. When the sphere colliders of two separate s orbital objects intersect, the bonding or anti-bonding molecular orbital is generated depending on the phase of the atomic orbitals (which can be toggled by the user). In figure 6.10, the anti-bonding molecular orbital is also presented, displaying a further piece of functionality, the capability to toggle the nodal planes (if present).

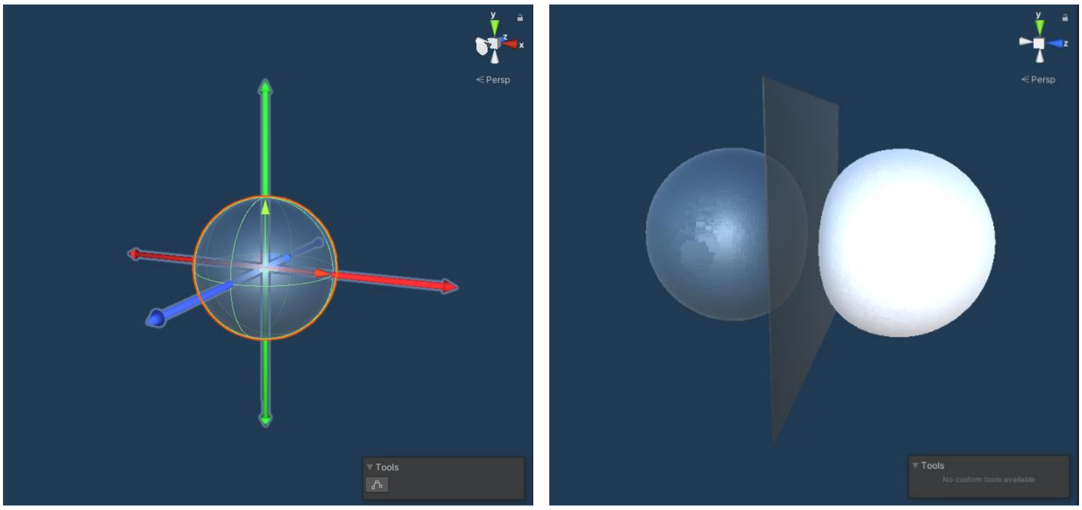



Figure 6.10. A 1s orbital with superimposed axis and a sphere collider (left), and a s-s antibonding molecular orbital (right) with toggled nodal plane.

The development of the 2p orbitals introduced a level of complexity not observed when developing the s orbitals, the issue of symmetry. Unlike the s orbitals, the direction of approach, and how the p orbitals intersected with other atomic orbitals influences the bonding interaction. To overcome this, two sets of colliders (with separate C# components) were attached to the 2p atomic orbital virtual objects (figure 6.11).

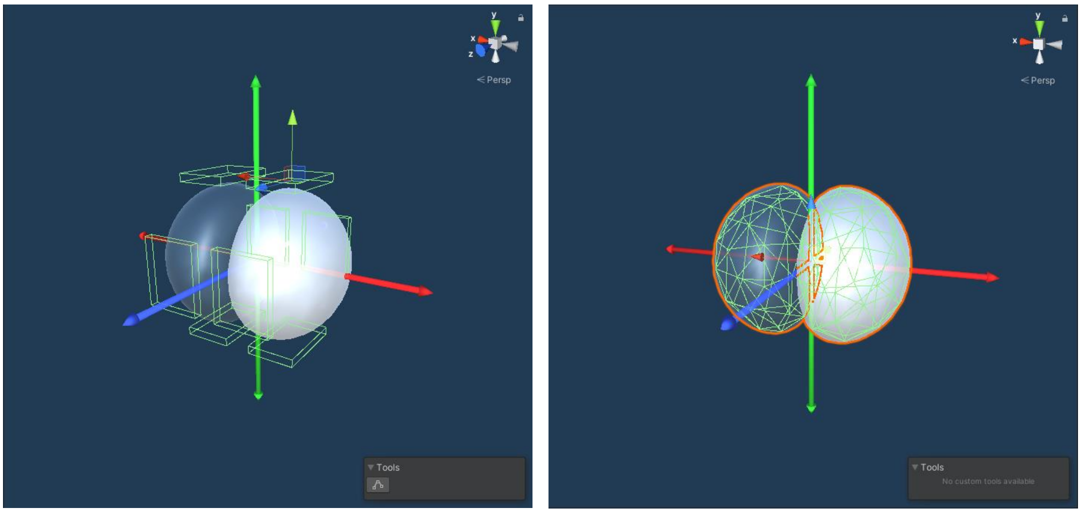



Figure 6.11. Box colliders controlling the pi interactions (left) and the mesh colliders controlling the sigma interactions (right).

The same level of complexity was also present when incorporating functionality into the d orbitals, to control for pi (π) and delta interactions (δ). Figure 6.12 presents the two sets of colliders incorporated into d orbital virtual objects. The mesh colliders wrapped directly around each lobe control the π bonding interactions, whereas box colliders situated above and below control the δ bonding interactions. On intersection, the OnCollisionEnter() method is called. An example of the condensed version of the delta interaction is shown in figure 6.13. Line numbers are provided on the left.

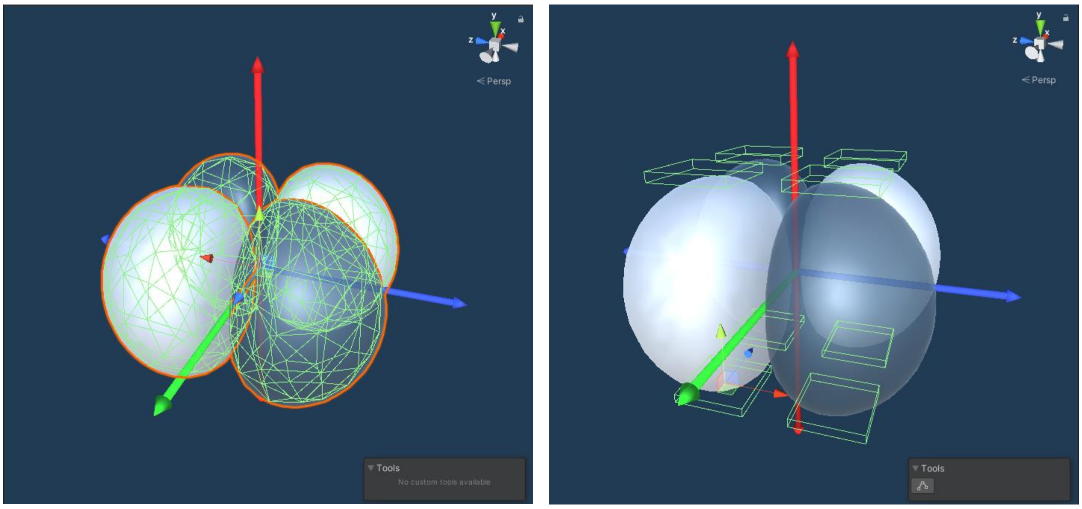



Figure 6.12. Mesh colliders controlling the pi interactions (left) and the box colliders controlling the delta interactions (right).

```
// Method called when collider conditions are met for a delta interaction
Unity Message | 0 references
private void OnCollisionEnter(Collision col)
{
    // Which direction did the orbitals approach from?
    Vector3 _direction = col.contacts[0].point - transform.position;
    _direction = _direction.normalized;

    // Check that the colliders don't share the same parent gameobject (otherwise they will form the MO on spawn).
    if (gameObject.transform.parent.GetInstanceID() != col.gameObject.transform.parent.GetInstanceID())
    {
        // Check the phase of the orbital lobes.
        if (col.gameObject.CompareTag("Delta Collider") && gameObject.CompareTag("Delta Collider") ||
          col.gameObject.CompareTag("Delta Collider OoP") && gameObject.CompareTag("Delta Collider OoP"))...
    }
    // Check that the colliders don't share the same parent gameobject (otherwise they will form the MO on spawn)
    if (gameObject.transform.parent.GetInstanceID() != col.gameObject.transform.parent.GetInstanceID())
    {
        // Check the phase of the orbital lobes.
        if (col.gameObject.CompareTag("Delta Collider") && gameObject.CompareTag("Delta Collider OoP") ||
          col.gameObject.CompareTag("Delta Collider OoP") && gameObject.CompareTag("Delta Collider"))...
    }
}

// Morph the atomic orbitals into a bonding or antibonding MO
2 references
private void GenerateOrbital(Vector3 point, Transform orbitalObject)
{
    Transform MOTransform = orbitalObject;
    MOTransform.position = point;
    MOTransform.localScale = new Vector3(transform.localScale.x * 0.75f, transform.localScale.x * 0.75f,
      transform.localScale.x * 0.75f);
}
}
```

Figure 6. 13. The OnCollisionEnter() method for delta bonding interactions.

## 6.6 Qualitative Data Analysis

Qualitative analysis of students’ discussions was completed through latent thematic analysis using the approach of Braun and Clarke (2006). All collected data was transcribed verbatim, prior to being subjected to analysis for commonly occurring themes. For this thematic analysis, themes were constructed around identified evidence of resource activation, as defined by Tuminaro and Redish (2007). The critical elements of this model are the basic elements of knowledge stored in long-term memory, the way those elements are linked, and the way in which those lined structures are activated in different circumstances (Tuminaro and Redish, 2007).

Three resource types are described. The first are *knowledge elements* used to describe irreducible cognitive structures held in long-term memory. Here, a knowledge element refers to a piece of chemistry information that students use within their discussions. Secondly, *knowledge structures* are used to link patterns of association between knowledge elements. The linking of existing knowledge is a basic assumption of Constructivism. Scott, Mortimer, and Ametller (2011) identify three forms of pedagogical link-making:

i. **Support knowledge building.** Making connections between diverse kinds of knowledge to support students in developing a deep understanding of subject matter.
ii. **Supporting continuity.** Making references to teaching and learning activities across points in time.
iii. **Encouraging emotional engagement.** The diverse ways in which the instructor makes links to encourage a positive emotional response from students to the ongoing teaching and learning.

Lastly, *control structures* are used to determine when knowledge elements are activated. The control systems that are considered here are *epistemic games* (Tuminaro and Redish, 2007). Control structures differ from knowledge elements and knowledge structures as they are tacit, normally being activated at a subconscious level. Thus, students are often unaware that they are engaging in a particular epistemic game. However, this can be deduced from my recorded student discussions. As such, the focus of this qualitative analysis is the interaction between students' AR experiences and the activation of these three resource types.

### 6.6.1 Epistemic Games

To evaluate and understand students' experiences during peer discussion, descriptions are needed which analyse the way that resources are organised. Tuminaro and Redish (2007) describe six coherent organisational control structures based on epistemic games introduced by Collins and Ferguson (1993). In their work *Epistemic Forms and Epistemic Games*, Collins and Ferguson define an epistemic game as a complex "set of rules and strategies that define inquiry." These games are described as epistemic as students engage in them as a means of constructing knowledge. As such, epistemic games can be used to describe scientific inquiry. Yet, epistemic games are not confined to just students, but everyone. For example, if an individual is comparing two objects or ideas, one approach is to list the characteristics of each. This is regarded as the simplest compare-and-contrast game (Collins and Ferguson, 1993).

However, it is important to note that students are not likely participating in these games consciously and are even more unlikely to be capable of articulating the games being played. When we use epistemic games, we are describing the behaviour of students during peer instruction when using AR technology, not the student's knowledge of their own behaviour. Expanding on the compare-and-contrast game, every list is implicitly the answer to a question in that it constructs knowledge to satisfy a goal. Considering coordination chemistry, some examples may be: “What are the different d atomic orbitals?”, and “What are the different geometries a metal complex can adopt?”.

An epistemic game is composed of two ontological components:

i. A knowledge base.
ii. An epistemic form.

An epistemic game is an activation of a pattern of activities that can be associated with a collection of resources (Tuminaro and Redish, 2007). The collection of resources that a student draws upon when playing an epistemic game constitutes the knowledge base. Drawing on a previous example, to answer a question pertaining to the shapes adopted by metal complexes, an individual requires prerequisite knowledge of the valence shell electron pair repulsion theory (VSEPR), in addition to ligand field theory (LFT). The epistemic form is often denoted by an external representation, that helps to guide students’ inquiry. Referring to the compare-and-contrast epistemic game outlined by Collins and Ferguson (1993), the epistemic form is the list itself, guiding the progression of the inquiry.

Structurally, an epistemic game is composed predefined moves. These are the procedures that occur in the game and are contained within a set of entry and ending conditions. When solving chemistry problems, students’ expectations about chemistry problems determine the entry and ending conditions, which affects the strategy they employ. These preconceived epistemological stances will inevitably influence the epistemic game students choose to engage in. The critical element of an epistemic game is that each game specifies a certain set of moves. Throughout my analysis, evidence of two epistemic games, described by (Tuminaro and Redish, 2007) was apparent. These were the Pictorial Analysis and Recursive Plug-and-Chug epistemic games.

In **Pictorial Analysis**, students generate an external spatial representation that specifies the relationship between influences in a problem statement (Tuminaro and Redish, 2007). Within the AR environment, a student could generate a specific virtual object, which would act as an epistemic form to guide students’ inquiry. The moves in this game are determined by the external representation that the students choose. Despite differences that may arise based on the external representation chosen, there are moves that are common to all instantiations of the Pictorial Analysis Game (figure 6.14): a determination of the target concept followed by a choice of representation. Next, a student tells a conceptual story about the question posed based on the spatial relation between objects. Lastly, the slots of the representation are filled.


Figure 6.14. Schematic diagram of some moves in the epistemic game Pictorial Analysis. Adapted from Tuminaro and Redish (2007).

Secondly, in the **Recursive Plug-and-Chug** epistemic game, students plug ideas into a problem situation and churn out answers without conceptually understanding the implications of their solution. Students do not generally draw on their intuitive knowledge base while playing this game. Consequently, students engaging in Recursive Plug-and-Chug rely only on their syntactic understanding, without attempting to understand chemistry conceptually. In other words, other cognitive resources are usually inactive during this game (figure 6.15).

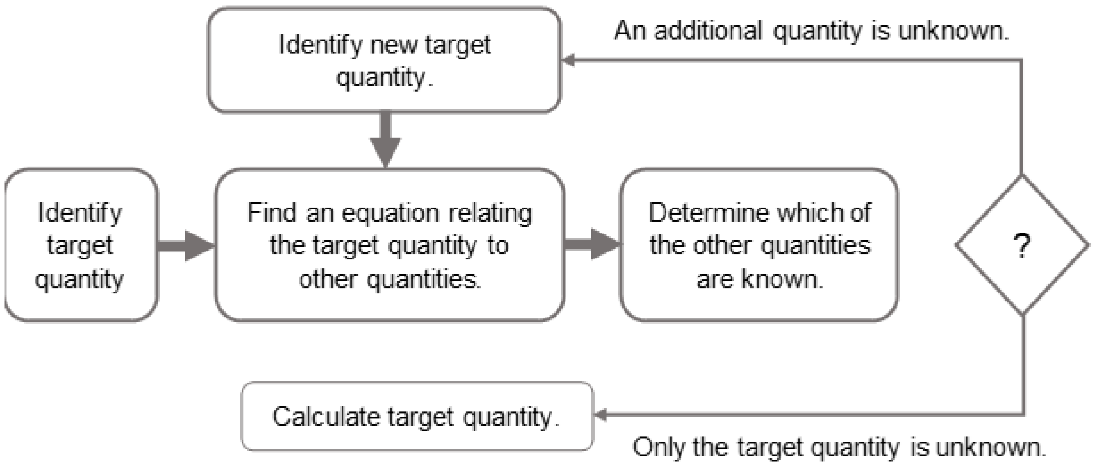



Figure 6.15. Schematic diagram of some moves in the epistemic game Recursive Plug-and-Chug. Adapted from Tuminaro and Redish (2007).

### 6.6.3 Evidence of Pictorial Analysis

I start by examining students' discussions for CTs 2 and 3, as these both showed significant intragroup improvement and high PI efficiency. CT2 relates to the identification of a linear complex's crystal field splitting diagram (see figure 6.3), whereas CT3 concerns the geometric [Jahn-Teller] distortion of a non-linear molecular system (see figure 6.4). These CTs provide examples of productive student dialogue in which a change in student thinking, and voting response, are evident. Upon inspection of students' responses in round 1, the majority voted for answers B (58%) and A (38%) for CTs 2 and 3, respectively. Throughout round 2, evidence of the Pictorial Analysis epistemic game was apparent throughout students' discussions (figure 6.16). The first discussion presented was between a pair of students, of which one voted correctly during round 1, and the other incorrectly. The second comment from group member (GM) 1 is the first activating statement in this dialogue. GM2 explains the interaction between the d orbitals of the gold atom, and the two chlorido ligands.

Amidst choosing a new d atomic orbital on the generated virtual object, there has been a change in thinking for GM1. This can be interpreted as an activating event, and evidence of the lowest level of resource activating, activation of a knowledge element. As the dialogue progresses, GM1 has understood the concept, and is now able to use their knowledge to contribute to the discussion. Combining the video recording representing the students' AR experience, with the audio recording of the peer discussion, gave a clear indication of the positive impact that using AR had on supporting students' thinking and knowledge construction. Both the voting statistics, and investigation of the dialogue, demonstrated a sound understanding of this CT across my cohort.

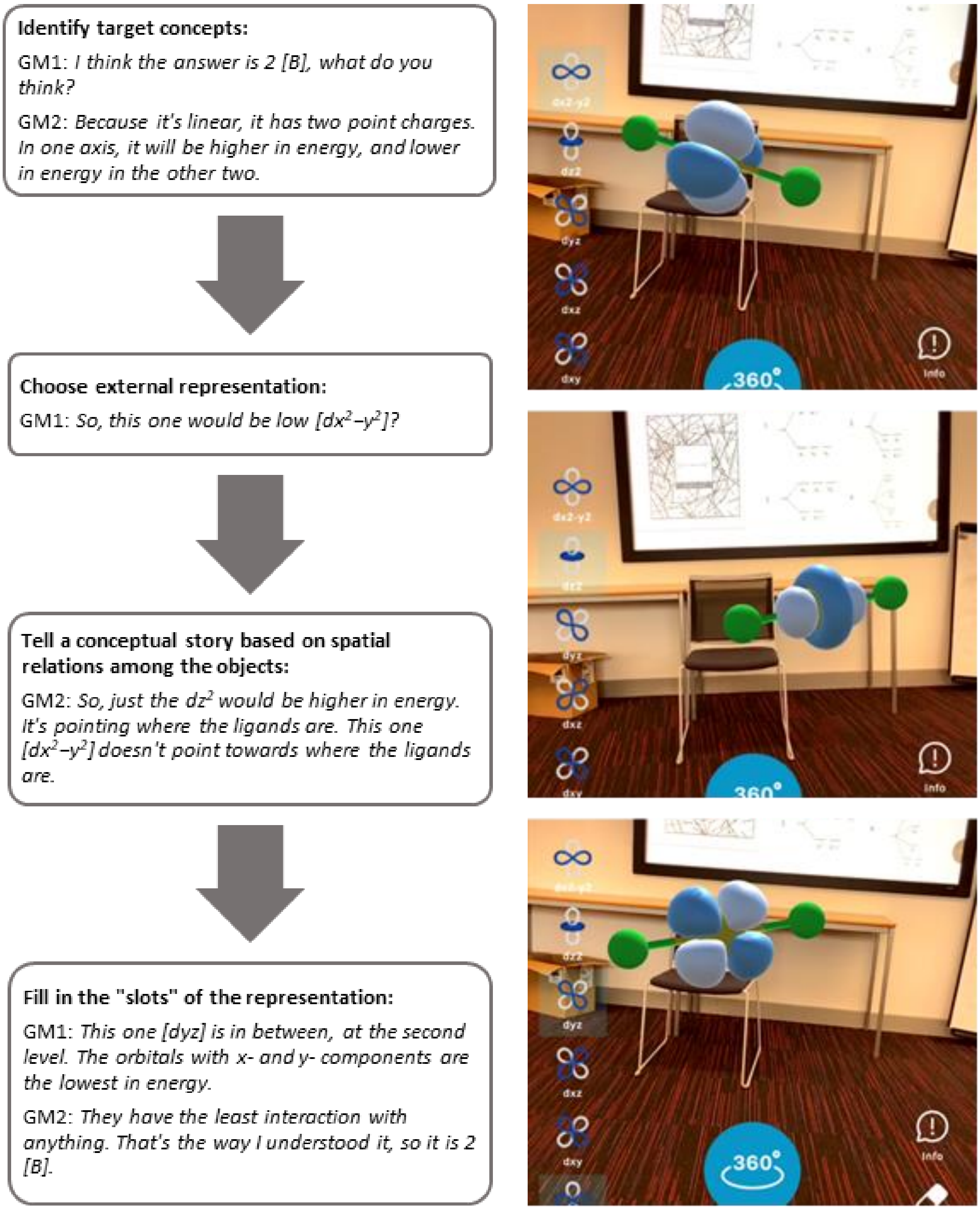



Figure 6.16. Students' discussions (left) alongside the corresponding AR experience (right) for CT2, showing some moves in the epistemic game Pictorial Analysis.

As the session progressed to CT3, employment of the Pictorial Analysis epistemic game was, again, evident from students' discussions. The example outlined in figure 6.17 is a group of three students, in which a single member answered correctly during round 1, and the other two incorrectly. The interplay that is of particular interest is section 2. GM2 can warrant proof of their claim using the provided AR tool. The distortion of the represented octahedral complex is used as a means of activating the thinking of GM3. GM3 demonstrates activation of a knowledge structure, specifically, support knowledge building. The thinking of GM1 has not changed. Thus, GM2, building on their previous statement, thinks of a new way to persuade GM1 regarding the stabilisation of the z-components. Using the AR tool, GM2 can introduce the metal d orbitals to support their conceptual story. Subsequently, GM1's thinking is activated, and repeats the statement that altered their perspective, reaching the correct conclusion. Although the two dialogues presented in figures 6.16 and 6.17 are short, it demonstrates many of the attributes that were commonly found in discussions. Both CTs provide evidence of resource activation by means of AR-supported dialogue. All three students responded correctly on the second round of voting.

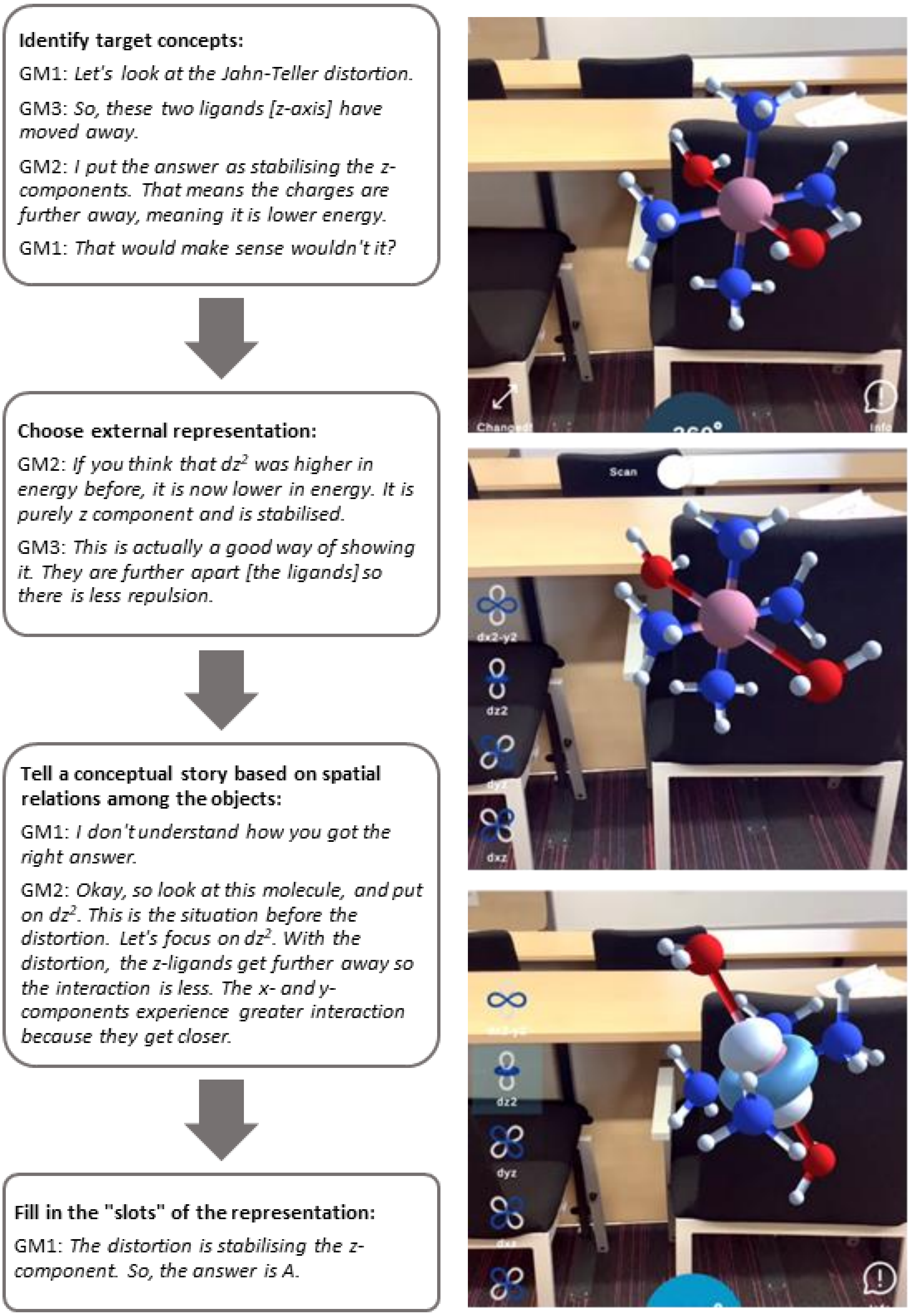



Figure 6.17. Students' discussions (left) alongside the corresponding AR experience (right) for CT2, showing some moves in the epistemic game Pictorial Analysis.

### 6.6.3 Evidence of Recursive Plug-and-Chug

CT1 is an interesting case. Although quantitative response data suggests that most students answered correctly, both before and after discussion, qualitative data suggests that students may not have demonstrated a clear understanding at the start of the dialogue. As such, there are points of interest in terms of resource activation through utilisation of the AR tools. Below, I present an example of a dialogue from a pair of students for CT1. The AR representations employed are shown in figure 6.18. Of note, is that both students answered correctly before and after peer discussion:

GM1: *I put D; I don't know.*

GM2: *So, if we look at the $dz^2$ and the $dx^2{-}y^2$, when it splits there will be 2 orbitals at the top and 3 on the bottom.*

GM1: *Yeah, but the question is why those ones?*

GM2: *It's this [$dz^2$] because the orbital is pointing towards the ligands. If you think of ligands as being point charges, the orbital overlaps with the ligands. That's higher energy. And this one also [$dx^2{-}y^2$]. The ones between the axis are in the $t_{2g}$.*

GM1: *It makes sense that the top two orbitals are in line with the ligands, that these ones [$dx^2{-}y^2$] are pointing towards the ligands which is unfavourable, so it's going to be the highest energy. These ones [dxy] are between the axis and therefore lower in energy.*

GM2: *These aren't pointing at the ligands, so I think that these are energetically favourable.*

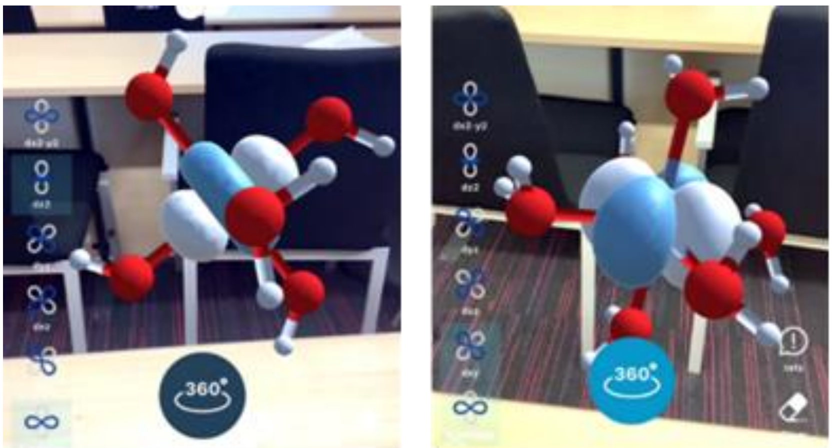


Figure 6.18. AR representations employed during peer discussion of CT1, with overlay of the $dz^2$ orbital (left), and the overlay of the dxy orbital (right).

As observed previously for dialogues concerning CT2 and CT3, students were able to use the AR tool to generate evidence to warrant their claims throughout discussion. GM1 further benefited via the contribution of GM2, who used to AR tool to not only activate knowledge elements, but to also link these elements into knowledge structures. This again demonstrates engagement with the Pictorial Analysis epistemic game, as well as a successful dialogue where students engaged in a productive discussion, where it is evident that a clear progress in understanding has been achieved.

Lastly, I provide an example from CT5, which required students to use their understanding of pi backbonding to identify which carbonyl ligands (figure 6.19) are most susceptible to electrophilic attack. My quantitative data shows that CT5 had the lowest correct response rate after independent voting, as well as the lowest theoretical (and measured) PI efficiency. Furthermore, it was the only situation where the correct response rate of students was lower after discussion. Hence, it is important to understand the interactions present throughout discussions of CT5, and how these differ from the successful dialogues presented in CTs 1−3.

In all the transcripts, a common theme was whether students could recognise that the two bridging carbonyl ligands are equivalent. Cognitively, my intention was that the 3D perspective afforded by ChemFord would not only help manage working memory load, but also afford students the opportunity to better observe the spatial relations between different ligands.

[Group 10]

GM2: *I'm thinking about sterics for sure. The top one will be sterically hindered.*

GM1: *The top one is next to two rings so that's not feasible. It's not going to be all of them. It's not going to be the two terminal ones.*

[Group 19]

GM1: *It's definitely not the top bridging CO.*

GM3: *The top one will be sterically hindered by the two other ligands.*

GM2: *Looking at the molecule you can see that both of the bridging COs are equivalent.*

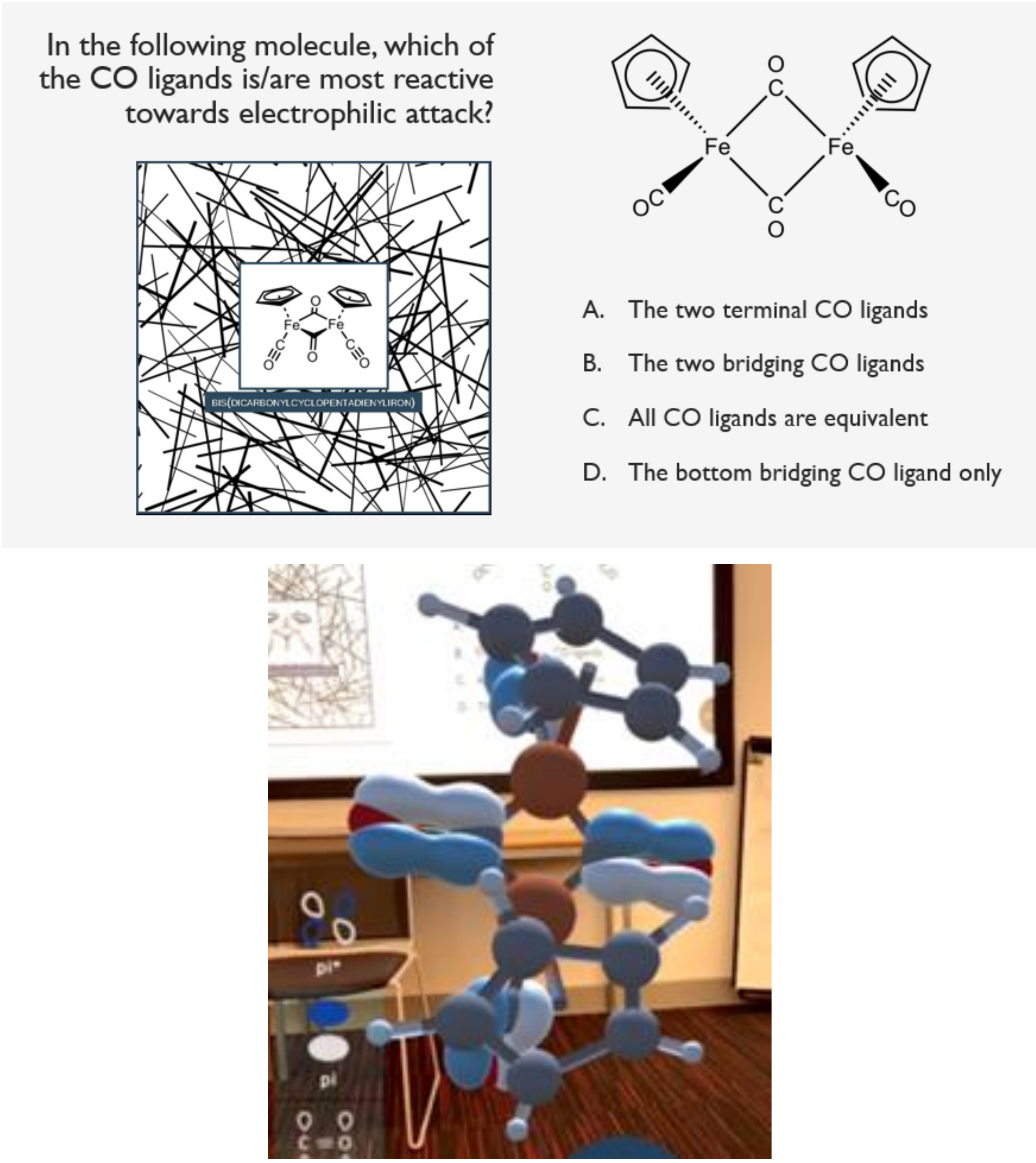



Figure 6.19. My fifth CT (top) and a 3D representation of cyclopentadienyliron dicarbonyl dimer with superimposed carbonyl π bonding molecular orbitals (bottom).

For this CT, I developed virtual representations of the π and π* molecular orbitals of the carbon monoxide ligands, in addition to the iron atom d orbitals, in the hope of initiating discussion of electron backdonation. This was noted in some dialogues, in which students responded correctly during round 2:

[Group 2]

GM1: *I'm just thinking about the antibonding orbitals on the carbonyls. The antibonding for these ones [carbonyls] would be here.*

GM2: *Okay.*

GM1: *And these ones are bound to two metals and these ones are only bound to one metal.*

[Group 5]

GM1: *Yeah. Backbonding also provides electrons to the ligand.*

GM2: *Yeah.*

GM1: *So, those go into the pi* orbital of the CO.*

However, several dialogues for this CT provided examples of unproductive discussion in which little conceptual chemistry were used in the dialogues. A reason for this may be that students were not able to retrieve the required knowledge elements to respond correctly, or that my AR experience did not manage to support resource activation. Evidence of dialogue like that expected of Recursive Plug-and-Chug epistemic game was also observed in CTs 4 and 6 (figure 6.20), but not in CTs 1−3. For group dialogues where the AR virtual objects were not referenced, or a driver for supporting the discussion, I found a greater number of incorrect responses after round 2.

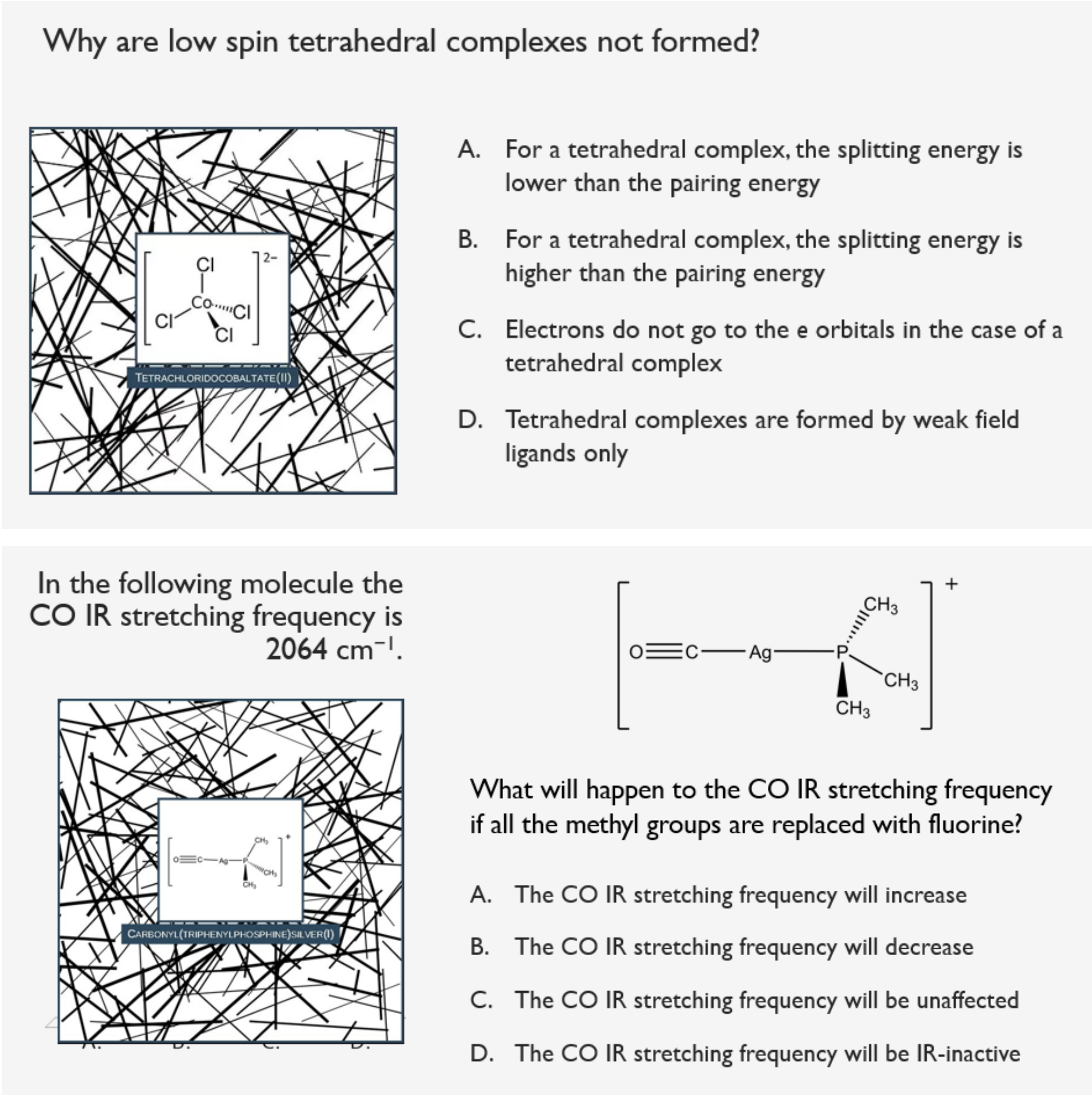



Figure 6.20. The fourth CT (*top*) and sixth CT (*bottom*) utilised in my PI session.

## 6.7 ConcepTest difficulty

Smith et al. (2009) report that students improve the most when instructors ask tough questions when implementing PI, a study that was replicated by Porter et al. (2011), in two computer science courses, who found the same trend. In addition, lower learning gains have also been reported for instructors implementing easier CTs (Rao and DiCarlo, 2000). Hence, empirical evidence suggests that the benefits of PI, especially the effectiveness of student discussions, is likely influenced by the difficulty of the question posed. In their longitudinal analysis, Crouch and Mazur (2001) found that the greatest learning gains following voting in round 2 occurred when the voting in round 1 was correct for ~50% of the student base. Yet, there were still substantial learning gains when the initial proportion of correct responses was between 35 and 70%. Below 35%, the concept may still be too alien, requiring the provision of further description. However, this does not mean that students are unable to benefit from peer discussion at lower levels of correct responses on round 1 (Simon et al., 2010). As per the implementation model of PI presented in figure 6.1, facilitators commonly provide only a brief explanation of the answer for correct response rates above 70%.

A crucial component of evaluating CTs is establishing reliability, thus providing users with information regarding the quality of each question. For this study, I applied the analytical procedures of IRT to calculate values of difficulty and discrimination for each CT. The fit of different parameter models to my data was evaluated using Orlando and Thissen's (2003) S-$\chi 2$ item-fit statistics (table 6.3), in addition to computed values of Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) fit statistics (table 6.4, Acquah, 2010).

Table 6.3. Item-fit statistics for 1PL, 2PL and 3PL IRT models. Statistically significant items are shown in blue.

| PI CT number | 1PL | | 2PL | | 3PL | |
|---|---|---|---|---|---|---|
| | ***S-χ2*** | ***p*** | ***S-χ2*** | ***p*** | ***S-χ2*** | ***p*** |
| 1 | 10.70 | 0.005 | 5.12 | 0.024 | 15.12 | NaN |
| 2 | 0.23 | 0.892 | 0.33 | 0.565 | 1.25 | NaN |
| 3 | 1.06 | 0.589 | 0.61 | 0.435 | 0.81 | NaN |
| 4 | 0.95 | 0.622 | 0.96 | 0.326 | 1.83 | NaN |
| 5 | 5.54 | 0.625 | 8.80 | 0.003 | 5.65 | NaN |
| 6 | 4.14 | 0.126 | 3.55 | 0.060 | 3.71 | NaN |

Table 6.4. IRT coefficients (2PL) for the six developed CTs.

| Model | Log-Likelihood | AIC | BIC |
|---|---|---|---|
| 2PL | −89.491 | 202.98 | 221.30 |
| 3PL | −84.466 | 204.93 | 232.41 |

My data did not show an improved model fit on addition of the pseudo-guessing parameter (3PL). Thus, I employed a two-parameter model (2PL) for my evaluation. The developed CTs demonstrate reasonable difficulty and discrimination (table 6.5), constituting evidence that each of the CTs is positively related to an individual's overall proficiency in this subject topic.

Table 6.5. IRT coefficients (2PL) for the six developed CTs.

| ConcepTest No. | Difficulty | Discrimination |
|---|---|---|
| 1 | −2.000 | 0.636 |
| 2 | −0.225 | 0.543 |
| 3 | 0.917 | 0.956 |
| 4 | 1.414 | 0.561 |
| 5 | 3.004 | 0.350 |
| 6 | 1.056 | 1.580 |

The item-characteristic curves (ICC) are shown in figure 6.21. Of the six CTs posed, 1 and 2 are considered the easiest, at the lower estimate of individuals' ability. CTs 3, 4 and 6 demonstrate difficulty values around the mean of the population distribution of the latent trait. CT5 is considered the hardest item, at the higher estimate of individuals' ability.

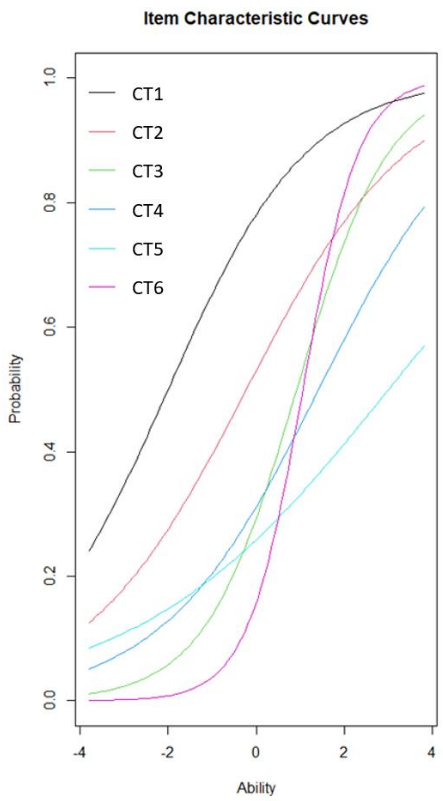



Figure 6.21. ICC generated from my 2PL IRT model for the 6 CTs posed.

I employed PI efficiency ($\eta$) calculations, defined with the help of Hake's standardised gain (Nitta, 2010), to further examine the effectiveness of each CT. The proportion of correct answers before, and after, the discussion is denoted by $N_b$ and *Na,* respectively. While Hake's gain represents individual learning gain, PI efficiency is considered to reflect the ease of understanding gained through PI (table 6.6). The collected response data from my CTs was found to be normally distributed. Hence, I conducted paired-samples *t*-tests, alongside analysis of effect size, for intragroup comparisons. The theoretical value of $N_a$ is expressed as a function of $N_b$ (Nitta, 2010), with the theoretical value of $\eta = N_b$. For this study, the average difference between the measured, and theoretical values of $\eta = 0.061$, similar to a value of 0.062 recorded by (Nitta, Matsuura and Kudo, 2014) when measuring the effectiveness of PI using the Force Concept Inventory. The proportion of correct responses during independent voting in round 1 ranged from 0.290−0.897, lying outside of the ideal range reported by Crouch and Mazur (2001). For CTs 4−6, where correct independent response rates lie at the lower end of this range, students were likely to have had ineffective discussions during round 2. As such, the value of $\eta$ observed is low.

Table 6.6. Correct answer proportion and PI efficiency of the CTs.

| **Measure** | **ConcepTest Number** | | | | | |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** |
| No. of respondents before PI discussion | 29 | 31 | 29 | 33 | 31 | 27 |
| No. of respondents after PI discussion | 32 | 31 | 29 | 25 | 30 | 25 |
| Correct answers before discussion ($N_b$) | 0.897 | 0.581 | 0.379 | 0.333 | 0.290 | 0.296 |
| Correct answers after discussion ($N_a$) | 0.969 | 0.968 | 0.862 | 0.400 | 0.200 | 0.320 |
| Paired Samples *t*-test | 0.161 | <0.01 | <0.01 | 0.161 | 0.375 | 1.000 |
| Cohen's d* | 0.28 | 0.77 | 0.91 | 0.29 | 0.17 | 0.00 |
| Theoretical value of $N_a$ | 0.989 | 0.824 | 0.614 | 0.555 | 0.496 | 0.504 |
| PI efficiency ($\eta$) | 0.699 | 0.924 | 0.778 | 0.100 | −0.127 | 0.034 |
| Theoretical value of PI efficiency ($\eta$) | 0.897 | 0.581 | 0.379 | 0.333 | 0.290 | 0.296 |
| Difference between theoretical and measured values | 0.198 | −0.343 | −0.399 | 0.233 | 0.417 | 0.262 |

* Limits for measures of Cohen's d: low ($c < 0.2$), medium ($0.2 \leq c < 0.5$), high ($0.5 \leq c$)

The normalised proportion of correct responses before, and after, the discussion phase of each CT is shown in figure 6.22. I observed statistically significant improvement for correct response rates between the first and second round of voting on CTs 2 and 3. For CTs 1 and 4, this improvement was approaching significance, with the difference between groups greater than 0.2 standard deviations.

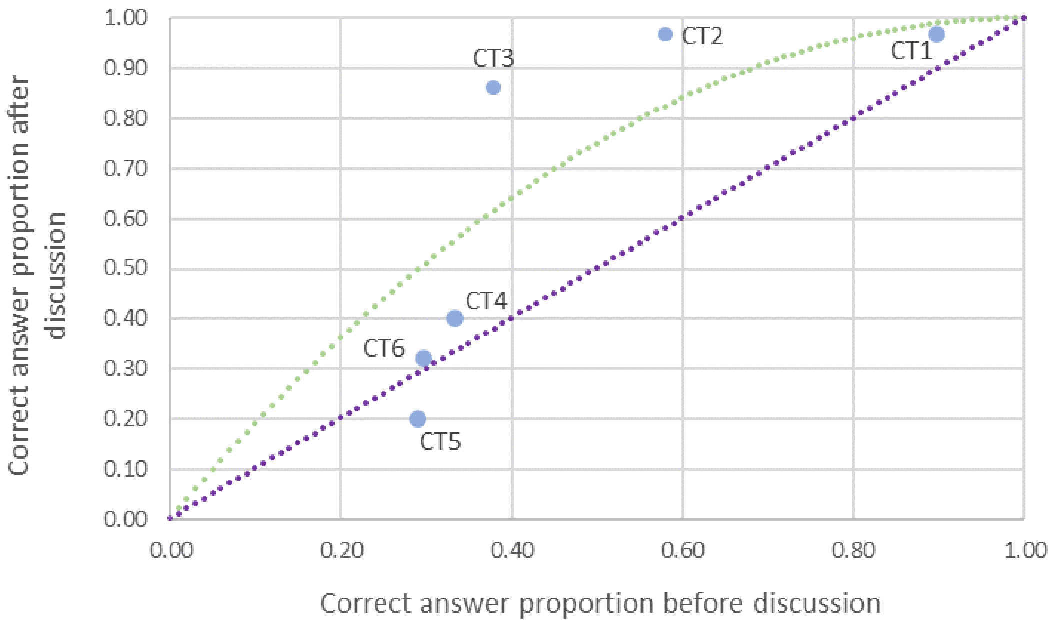



Figure 6.22. The proportion of correct responses before, and after, discussion of each CT. The green line represents the theoretical curve for PI efficiency (Nitta, 2010). The purple line represents equal pre-discussion and post-discussion accuracy. Points above this line indicate improvements in accuracy, whereas points below the line represent decrements in accuracy.

## 6.8 ConcepTest Response Switching

The descriptive statistics outlining the extent to which students switched their responses to each CT, between the first and second round of voting, are shown in table 6.7. Details pertaining to the proportion of students who switched in any direction, in addition to the proportion of responses that are switched in a specific direction [wrong-to-right (W-R); wrong-to-wrong (W-W); and right-to-wrong (R-W)] are provided. Throughout my PI session, the results of round 1 voting were shared with the cohort prior to round 2.

Another variation in the implementation of PI is the decision to display or describe the results of the first round of voting. This is commonly achieved by projecting a histogram or by describing the votes qualitatively. This may inevitably introduce unnecessary noise in the form of “the most common response bias” (and subsequently switching) between the first and second round of voting. Perez et al. (2010) report that seeing “the most common response can bias a student’s second vote on a CT and may be misinterpreted as an increase in performance due to student discussion alone”. This bias was more pronounced on tough questions, and it appeared to account for 5% of the learning gains observed between the first and second vote. Group interviews revealed that students perceive the most common answer to be the most correct, and students are less willing to defend an answer if it is not the most common one (Nielsen et al., 2012). Further work, reported by Brooks and Koretsky (2011), found that students shown the histogram after the first vote were statistically more confident when their answer matched the consensus answer, even if the consensus answer was incorrect. As such, more research is needed to fully understand the effect of displaying the histogram following the first round of voting.

Based on the results of the few studies investigating this issue, it may be most effective to show the difference in the distribution of answers between the first and second vote after peer discussion. This approach would limit the bias toward the consensus answer observed in some studies, while not only enhancing the confidence of students who had the correct answer in the first vote but also maintaining the integrity of student discussion. In addition, when the results of an initial vote are evenly split between two or more answers, displaying the graph may be a valuable conversation starter.

Table 6.7. The proportion of students’ responses that were switched between the first and second round of voting.

| **ConcepTest** | **Students who switched (%)** | **Direction** | | |
|---|---|---|---|---|
| | | Wrong-to-Right (W-R) | Wrong-to-Wrong (W-W) | Right-to-Wrong (R-W) |
| 1 | 29.41 | 70.0 | 10.0 | 20.0 |
| 2 | 48.48 | 81.3 | 12.5 | 6.3 |
| 3 | 65.63 | 71.4 | 23.8 | 4.8 |
| 4 | 63.64 | 28.6 | 38.1 | 33.3 |
| 5 | 65.63 | 19.0 | 47.6 | 33.3 |
| 6 | 54.84 | 23.5 | 52.9 | 23.5 |

When switching is measured, it is important to ensure that the data is not confounded with the frequency of correct (or incorrect) responses in round 1 (Miller et al., 2015). Normalising my response data with respect to students’ answers in round 1 provides an adjusted measure of switching, independent of how many times a student was correct, or incorrect, in round 1. Coupling these normalised values with the output of my 2PL IRT model allows me to examine switching (in any direction or a specific direction) as a function of CT difficulty (figure 6.23). A Pearson's correlation showed a strong, positive correlation, $r = 0.910$, between response switching and CT difficulty which was statistically significant, $p = 0.012$. With increasing CT difficulty, students are more likely to switch their answers from right-to-wrong ($r = 0.754$, $p = 0.084$), and wrong-to-different wrong ($r = 0.829$, $p = 0.042$). In addition, students are less likely to switch their answers from wrong-to-right ($r = −0.771$, $p = 0.072$). A finding consistent with previous studies (Miller et al., 2015).

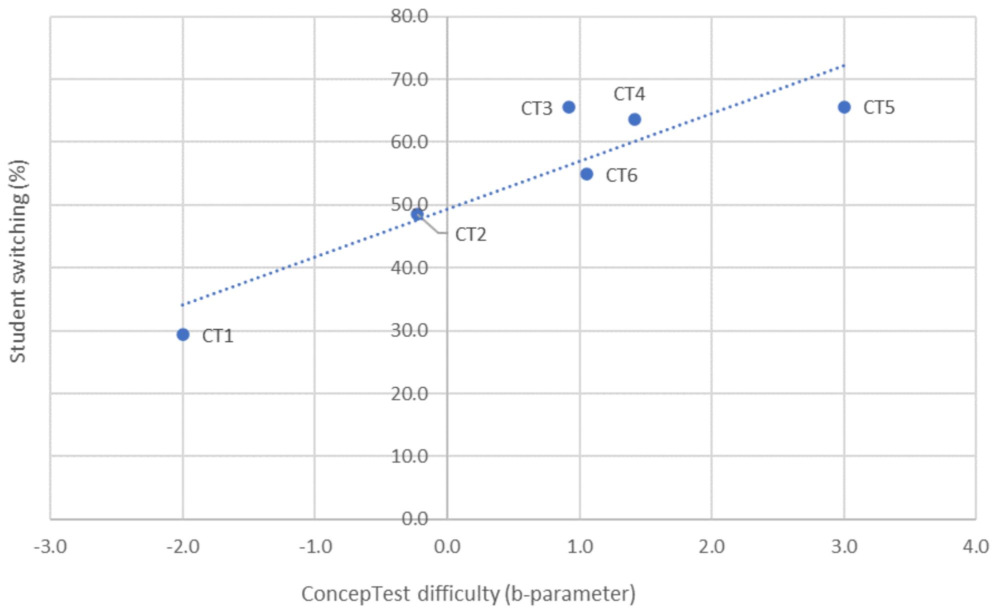



Figure 6.23. Student switching (%) in any direction for each CT as a function of difficulty. Each point represents a different CT.

It is important for instructors to understand that they have some control over the measure of response switching that occurs throughout PI via the difficulty of the CTs posed. Within my session, I attempted to scaffold this by posing easier CTs first, subsequently building up to more difficult CTs. Research has shown that prefacing more difficult problems with a sequence of related, but more basic conceptual questions, helps students answer harder problems (Ding, Reay, Lee, and Bao, 2011). Cognitively, presenting easier questions prior to tough questions may help students break down concepts into smaller, more manageable chunks. As CTs often require students to apply conceptual understanding in new contexts, it is possible that scaffolding difficult CTs may assist with positive switching transitions. A future study of CT response patterns to a series of scaffolded questions would prove interesting in providing further insight into the relationship between switching and CT difficulty.

## 6.9 Reported Self-Efficacy

From examination of students' reported measures, evidence of relationships between response switching and pre-session self-efficacy were found. Students reporting lower measures of self-efficacy were more likely to switch their responses in a negative direction (right-to-wrong and wrong-to-different wrong) in comparison to students with higher self-efficacy. Additionally, students reporting higher self-efficacy were more likely to switch from wrong-to-right than students with lower self-efficacy. Figure 6.24 shows the normalised proportion of switched responses for students with lower and higher self-efficacy. To analyse differences in self-efficacy, the cohort was divided into the top and bottom 27% (Preacher, 2015) based on students' reported measures from the PISE instrument.

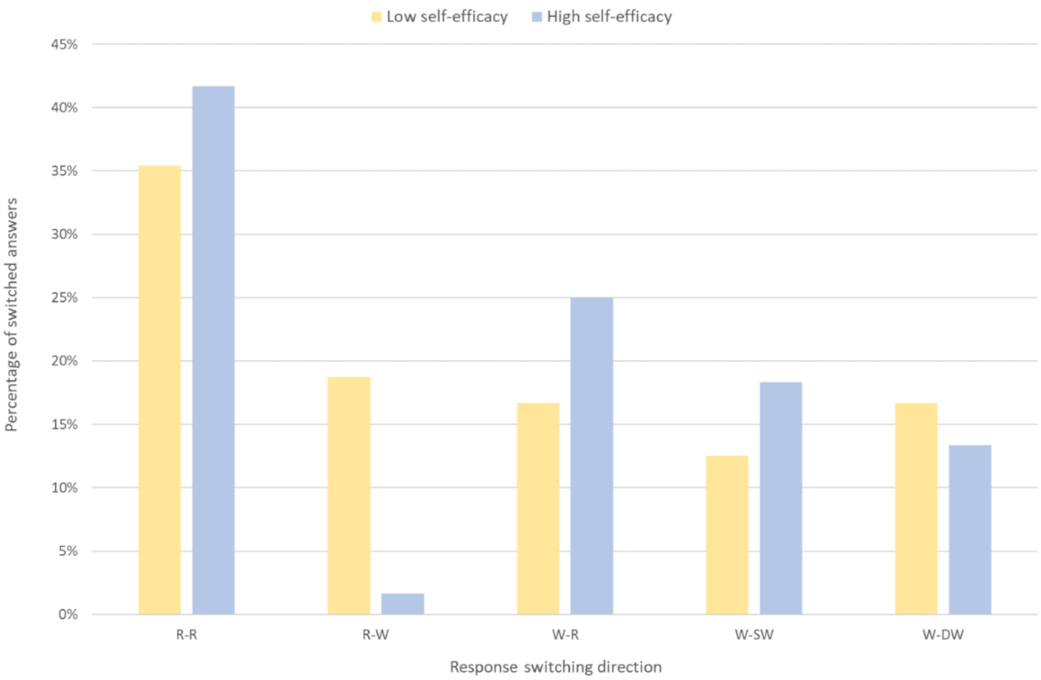



Figure 6. 24. Response switching patterns for students in the top and bottom 27% of reported self-efficacy measures.

Students with higher pre-session self-efficacy switched direction, regarding their responses, from right-to-wrong ($p < 0.05$) and wrong-to-different wrong less often;

and switched from wrong-to-right more often than students with lower self-efficacy. As I did not administer a pre-test assessment, I am unable to control for covariates such as prior knowledge, but previous work has indicated that self-efficacy may be more predictive of switching than incoming knowledge (Zajacova, Lynch and Espenshade, 2005). In addition, students with higher self-efficacy switched their CT responses from wrong-to-right more often than students with lower self-efficacy.

Students' responses to two individual items (10 and 16) on the PISE moderately correlated with switching from right-to-right. In other words, choosing the correct answer and then sticking with that response throughout round 2. These statements are: *"I usually don't worry about my ability to solve chemistry problems"*, ($p$ < 0.01); and *"I know how to explain my answers to organometallic chemistry questions in a way that helps others understand my answer"*, ($p$ < 0.01). In contrast, these two same items strongly negatively correlated ($p$ < 0.01) for switching from right-to-wrong answers. For item 10, students who either disagreed or strongly disagreed switched from right-to-wrong, regarding their CT response, significantly more than students who agreed or strongly agreed ($p$ < 0.001). This difference was also observed for item 16 ($p$ = 0.01). Students with a low assessment of their problem solving and science communication abilities are significantly more likely to switch their CT responses from right-to-wrong than students with a higher assessment of those abilities.

Following my PI session, median Likert scores on the PISE instrument improved on the following items:

*"When I come across a tough chemistry problem, I work at it until I solve it"* (neutral to agree).

*"I like hearing about questions that other students have about chemistry"* (neutral to agree).

*"I can communicate science effectively"* (neutral to agree, $p$ = 0.04).

*"I can communicate chemistry effectively"* (neutral to agree, $p$ = 0.025).

## 6.10 Study Limitations

The limitations of this study must be observed. Firstly, the PI evaluation is based on data gathered from one session and six CTs. It is therefore difficulty to generalise the results. Quantitative data concerning conceptual understanding was collected solely through CT voting data. Objective tests, such as the Force Concept Inventory, have been used previously to evaluate entire PI sessions (Hestenes, Wells and Swackhamer, 1992). In addition, the data analysis was based upon a relatively small sample size. This was the result of modest enrolment for my PI session.

## 6.11 Chapter Conclusions

This study contributes to the growing body of evidence on how students can benefit from the pedagogical approach of AR-supported PI. In summary, evidence of the interaction between CT difficulty and response switching has been found. As such, it is important for instructors to understand that they have a degree of control over the measure of response switching occurring throughout a PI session via the difficulty of the CTs posed. The difficulty values of the six developed CTs were calculated using a 2PL IRT model. The output of the 2PL model showed adequate difficulty and discrimination values for the CTs developed. In addition, the effectiveness of each CT was evaluated using PI efficiency calculations. The value of $\eta$ was highest in CTs 1–3.

Moreover, the relationship between response switching and reported self-efficacy was examined. Students reporting higher measures of self-efficacy displayed lower levels of switching in a negative direction. Students with a lower assessment of their problem solving and science communication abilities were significantly more likely to switch their responses from right-to-wrong than students with a high assessment of those abilities. Qualitative insights have provided evidence of epistemic games such as Pictorial Analysis within AR-supported PI discussions. Where calculated peer efficiency values for CTs were lower, this was less apparent. In these cases, Recursive Plug-and-Chug was the commonly observed control structure. Hence, the question evolves to how affordances of AR technology can be leveraged to subconsciously engage students in epistemic games such as Pictorial Analysis.

# 7

# Conclusion

The analysis presented in this thesis was formulated to address a significant gap in the literature; to understand the relationship between the utilisation of augmented reality (AR)-supported educational interventions and students' conceptual understanding of chemistry, over a timescale greater than those typically evaluated by cross-sectional studies. In response, four AR-supported educational interventions have been developed, with cognitive and affective factors, in addition to academic performance, evaluated. To enable a wide variety of augmented experiences, ChemFord was developed, utilising both marker-based and markerless approaches. The chosen demographic was undergraduate chemistry students enrolled at the University of East Anglia (UEA). The period of research was two academic years commencing in October 2020 and finishing in May 2022. A mixed-methods approach was employed comprising of quantitative and qualitative methods of data collection and analysis.

Regarding research question 1, the relevant chemistry experience of participants was assessed prior to their engagement with each of the developed interventions. For the educational escape activities (EEA) and worked-examples activity described in chapters 3 and 5 respectively, two instruments were developed as a means to assess students' conceptual understanding. In chapter 3, the creation of the Isomerism in Transition Metal Compounds (ITMC) instrument is discussed, with the development of the $S_EAr$ test instrument in chapter 5. A two-step validation approach was employed to ensure that the items on both instruments were appropriate to gauge students' learning gains. Firstly, internal validation was conducted where each consulted expert was asked to carefully read each item, and to see whether they agreed unambiguously with the selected answer, and to comment upon whether they agreed that the item was fit for purpose. Following internal validation, one round of external validation was carried out with experts from other UK-based universities. Following data collection, item and scale characteristics of both instruments were analysed using procedures of Classical Test Theory (CTT) and Item Response Theory (IRT) to gain insight into the overall instrument reliability. For the ITMC test instrument, items 1 and 6 were shown to be the easiest items on the scale. DIF showed no biased measurement of ability between groups when using the Raju Signed Area method. Regarding the $S_EAr$ instrument, items 1–4 are generally at the lower estimate of individuals' ability, whereas items 5–9 demonstrate difficulty values around the mean of the population distribution for ability. After each educational intervention, instruments were completed by participants again. This was done to allow learning gain calculations to be completed.

For the EEA introduced in chapter 3, the introduction of ChemFord, over and above the activity, did not result in any significant differences in post-test scores between the two experimental groups. This was also observed for the Game-Based Learning activity described in chapter 4. During the activity, participants from the AR group scored higher on submitted answers using my measurement rubric, and also stated that the integration of AR, as an additional mode of teaching, improved their understanding of VSEPR subject content. However, this was not reflected in the administered post-test. In fact, the control group was statistically better on items pertaining to molecular geometry. Similarly, as discussed in chapter 5, no significant differences were observed between groups for conceptual understanding when engaging with the worked examples activity, demonstrated by the mean scores achieved on the $S_EAr$ instrument. Yet, significant intragroup improvement and greater normalised change values were observed for the AR group. No significant intragroup improvement was found in the control group for conceptual understanding. Qualitative insights from the peer instruction session described in chapter 6 have provided evidence of epistemic games such as Pictorial Analysis within AR-supported PI discussions. Where this control structure was observed, students commonly submitted the correct answer as their response to the posed ConcepTest. Hence, the question evolves to how affordances of AR technology can be leveraged to subconsciously engage students in epistemic games such as Pictorial Analysis.

Common to all of the educational interventions that were developed throughout this thesis was the ability to critically identify and amend software errors in the AR environment, as this will cause confusion within the learning experience. As such, the significant technical expertise required to develop, include, and maintain interactive computer-based systems is still considered a major drawback of AR implementation within educational settings. For successful implementation, facilitators will require the skills to recognize and troubleshoot problems, whilst developers will need to carefully test applications, to ensure high educational value, and avoid unnecessary cognitive load. In summary, a positive opinion of all four educational interventions, and ChemFord, ran throughout most participants' discussions. For example, when discussing mental visualisation of structures, in relation to the topic of VSEPR, the role of AR in assisting the visualisation process was perceived to be of great benefit to students. However, no significant differences were detected between experimental groups for levels of cognitive load. It is suspected that difficulty stemming from the game mechanics confounded with the potential benefits of the AR technology. Yet, both experimental groups demonstrated significant improvements in spatial ability over the study period, with no significant differences observed in terms of gender performance. A moderate correlation was found between spatial ability and VSEPR test instrument performance. Again, intergroup comparisons did not show any significant differences between experimental groups.

Commenting further on research question 2, QCM measures of challenge, interest, and probability of success in chapter 5 (pertaining to achievement motivation) were found to correlate positively with reported germane cognitive load (GCL). Reported extraneous cognitive load (ECL) negatively correlated with reported GCL, in addition to measures of challenge and interest. Measures of challenge and interest demonstrated a stronger negative correlation with ECL for students displaying higher prior relevant chemistry experience. Participants displaying higher prior conceptual knowledge also reported higher measures of ECL, alongside lower normalised change values. As learner expertise increases, a shift to a heavier emphasis on problem solving may be beneficial. For learners with lower relevant chemistry experience, challenge was strongly correlated with probability of success.

The EEA described in chapter 3 further illustrates how design features of an AR-supported educational intervention can be employed to support motivation, and that is scalable to large student cohorts. Examples of extrinsic motivational factors were mentioned by interview respondents. Further work examining the effects of intrinsic and extrinsic motivation on students' perception and performance in educational escape initiatives would be a welcome addition. Through students' discussions, I have provided evidence of how design aspects of the EEA support the psychological need satisfaction outlined by SDT. This indicates how future evolution of the activity can address these needs. Reported measures of competency were seen as a positive predictor of intrinsic motivation. However, in this study, this was not observed to be a positive predictor of academic performance. Yet, significant intragroup academic improvement was observed in both experimental groups. Future work should concentrate on defining the game mechanics most appropriate to addressing both the pedagogical and learning objectives. There is a potential to gain considerable information about the development of key chemistry competencies using this teaching strategy. Research regarding how key elements of the EEA, such as the briefing and debriefing sessions, should be designed and constructed is yet to yield definitive insights. This further extends to the peer instruction session conducted in chapter 6. For students responding incorrectly, it would be beneficial to develop automated tailored feedback to ConcepTests, and to understand how this impacts learning.

Evidence of the interaction between ConcepTest difficulty and response switching has been found. Moreover, the relationship between response switching and reported self-efficacy was examined. Students reporting higher measures of self-efficacy displayed lower levels of switching in a negative direction. Students with a lower assessment of their problem solving and science communication abilities were significantly more likely to switch their responses from right-to-wrong than students with a high assessment of those abilities.

Qualitative data pertaining to research question 3 was captured through a series of semi-structured interviews with participants over the course of the research period. Interview schedules were constructed around the topics being investigated. Qualitative analysis of participants' interview responses was completed through latent thematic analysis using the approach of Braun and Clarke (2006). Negotiated agreement was employed to ensure reliability. The measure of agreement among coders was calculated using Krippendorff's alpha reliability coefficient. Commenting on chapter 5, qualitative data suggests how CTML design principles may have supported learning, as well as how participants conveyed their understanding of $S_EAr$ concepts following my intervention. Throughout all the semi-structured interviews conducted, students commented that the addition of AR to learning environments adds educational value, whether through mental visualisation support, or affording opportunities for collaborative learning. Of note, was the minority of students who expressed little to no previous experience using AR technology prior to the research period. Future research could embed the use of Cloud Anchors to create collaborative augmented experiences that could be shared by multiple people across many different devices (Google Developers, 2022). This would substantially enhance collaboration and accessibility within augmented environments, and is a direction in which the ChemFord application will be developed. I encourage this as an avenue of future research, which will only become more evident as devices get more sophisticated, and therefore capable of sustaining augmented reality experiences. Yet, the technical expertise required to incorporate this functionality would be substantial.

In addition, how AR technology is integrated into the learning environment, when considering physical space should also be considered. At present, many traditional higher education spaces are constructed to accommodate a didactic teaching approach, serving the purpose of a mass education. It is likely that the optimal learning environment, one that encourages collaboration and interaction with AR technology, removes the physical constraints of this spatial arrangement – the rows of seats and lack of working space.

To close, I have provided a quote from a student discussion which I believe encapsulates the work conducted throughout this research period:

*"We generally work in groups of three or four people and we don't get to do that very often. When you can talk with the other students, you can better understand how they think, how your ideas of the concept compare to their ideas. You can get a different view. And I enjoyed that because sometimes I have misconceptions of what is right, and then someone else says, 'Oh, you are only looking at it this way instead of that way'... And having augmented reality on devices, to look at how molecules are shaped in space, to support these discussions. That was memorable and unique."*

# Bibliography

Abd Majid, N.A. and Abd Majid, N., 2018. Augmented reality to promote guided discovery learning for STEM learning. *Int. J. on Advanced Science, Engineering and Information Technology*, *8*(4-2), pp.1494-1500.

Abdinejad, M., Talaie, B., Qorbani, H.S. and Dalili, S., 2021. Student perceptions using augmented reality and 3d visualization technologies in chemistry education. *Journal of Science Education and Technology*, 30(1), pp.87-96.

Abdollahi, A.M., Masento, N.A., Vepsäläinen, H., Mijal, M., Gromadzka, M. and Fogelholm, M., 2021. Investigating the Effectiveness of an Educational Escape Game for Increasing Nutrition-Related Knowledge in Young Adolescents: A Pilot Study. *Frontiers in nutrition*, 8, p.674404.

Acosta, J.L.B., Navarro, S.M.B., Gesa, R.F. and Kinshuk, K., 2019. Framework for designing motivational augmented reality applications in vocational education and training. *Australasian Journal of Educational Technology*, *35*(3).

Acquah, H., 2010. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics*. 2. 1-6.

Adams, W.K., Wieman, C.E., Perkins, K.K. and Barbera, J., 2008. Modifying and validating the Colorado Learning Attitudes about Science Survey for use in chemistry. *Journal of Chemical Education*, 85(10), p.1435.

AECT., 2014. *Handbook of research on educational communications and technology*. New York: Springer, pp.958.

Aiken, L.S., West, S.G. and Reno, R.R., 1991. *Multiple regression: Testing and interpreting interactions*. sage.

Ainsworth, S. and Th Loizou, A., 2003. The effects of self-explaining when learning with text or diagrams. *Cognitive science*, 27(4), pp.669-681.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), pp.716-723.

Akçayır, M. and Akçayır, G., 2017. Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, 20, pp.1-11.

Akçayır, M., Akçayır, G., Pektaş, H. and Ocak, M., 2016. Augmented reality in science laboratories: The effects of augmented reality on university students’ laboratory skills and attitudes toward science laboratories. *Computers in Human Behavior*, 57, pp.334-342.

Akiduki, H., Nishiike, S., Watanabe, H., Matsuoka, K., Kubo, T. and Takeda, N., 2003. Visual-vestibular conflict induced by virtual reality in humans. *Neuroscience letters*, 340(3), pp.197-200.

Albert, B. and Tullis, T., 2013. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. 2nd ed.

Almousa, O., Prates, J., Yeslam, N., Mac Gregor, D., Zhang, J., Phan, V., Nielsen, M., Smith, R. and Qayumi, K., 2019. Virtual Reality Simulation Technology for Cardiopulmonary Resuscitation Training: An Innovative Hybrid System With Haptic Feedback. *Simulation & Gaming*, 50(1), pp.6-22.

Alseadoon, I., Ahmad, A., Alkhalil, A. and Sultan, K., 2021. Migration of existing software systems to mobile computing platforms: a systematic mapping study. *Frontiers of Computer Science*, 15(2), pp.1-25.

An, J. and Holme, T., 2021. Evaluation of Augmented Reality Application Usage and Measuring Students' Attitudes toward Instrumentation. *Journal of Chemical Education*, 98(4), pp.1458-1464.

Anderson, J., Reder, L. and Simon, H., 1996. Situated Learning and Education. *Educational Researcher*, 25(4), pp.5-11.

Anderson, T. and Rivera-Vargas, P., 2020. A Critical look at Educational Technology from a Distance Education Perspective. *Digital Education Review*, (37), pp.208-229.

Andrew, S., 1998. Self-efficacy as a predictor of academic performance in science. *Journal of advanced nursing*, 27(3), pp.596-603.

Ang, J.W.J., Ng, Y.N.A. and Liew, R.S., 2020. Physical and digital educational escape room for teaching chemical bonding. *Journal of Chemical Education*, 97(9), pp.2849-2856.

Antonietti, A., Rasi, C., Imperio, E. and Sacco, M., 2000. The representation of virtual reality in education. *Education and Information Technologies*, *5*(4), pp.317-327.

Antonioli, M., Blake, C. and Sparks, K., 2014. Augmented reality applications in education. *The Journal of technology studies*, pp.96-107.

Arnab, S., Lim, T., Carvalho, M.B., Bellotti, F., De Freitas, S., Louchart, S., Suttie, N., Berta, R. and De Gloria, A., 2015. Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, 46(2), pp.391-411.

Arsenault, R. and Ware, C., 2004. The importance of stereo and eye-coupled perspective for eye-hand coordination in fish tank VR. *Presence*, 13(5), pp.549-559.

Atkinson, J.W., 1957. Motivational determinants of risk-taking behavior. *Psychological review*, 64(6p1), p.359.

Atkinson, R.K., Derry, S.J., Renkl, A. and Wortham, D., 2000. Learning from examples: Instructional principles from the worked examples research. *Review of educational research*, 70(2), pp.181-214.

Atkinson, R.K., Renkl, A. and Merrill, M.M., 2003. Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of educational psychology*, 95(4), p.774.

Avargil, S., Shwartz, G. and Zemel, Y., 2021. Educational Escape Room: Break Dalton's Code and Escape!. *Journal of Chemical Education*, 98(7), pp.2313-2322.

Aw, J.K., Boellaard, K.C., Tan, T.K., Yap, J., Loh, Y.P., Colasson, B., Blanc, E., Lam, Y. and Fung, F.M., 2020. Interacting with three-dimensional molecular structures using an augmented reality mobile app.

Awan, R.U.N., Noureen, G. and Naz, A., 2011. A Study of Relationship between Achievement Motivation, Self Concept and Achievement in English and Mathematics at Secondary Level. *International education studies*, 4(3), pp.72-79.

Ayres, P. and Sweller, J., 2005. The split-attention principle in multimedia learning. *The Cambridge handbook of multimedia learning*, *2*, pp.135-146.

Azuma, R.T., 1997. A survey of augmented reality. *Presence: teleoperators & virtual environments*, *6*(4), pp.355-385.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S. and MacIntyre, B., 2001. Recent advances in augmented reality. *IEEE computer graphics and applications*, 21(6), pp.34-47.

Azuma, R. and Bishop, G., 1994, July. Improving static and dynamic registration in an optical see-through HMD. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques* (pp. 197-204).

Bacca, J., Baldiris, S. and Fabregat, R., 2018. Insights into the factors influencing student motivation in augmented reality learning experiences in vocational education and training. *Frontiers in psychology*, p.1486.

Bacca-Acosta, J. Baldiris, S. Fabregat, R. Sabine, G. Kinshuk, Dr., 2014. Augmented Reality Trends in Education: A Systematic Review of Research and Applications. *Educational Technology and Society,* 17, pp.133-149.

Baddeley, A.D. and Lieberman, K., 2017. Spatial working memory. In *Exploring working memory* (pp. 206-223). Routledge.

Badilla-Quintana, M.G., Sepulveda-Valenzuela, E. and Salazar Arias, M., 2020. Augmented reality as a sustainable technology to improve academic achievement in students with and without special educational needs. *Sustainability*, 12(19), p.8116.

Bagozzi, R.P. and Burnkrant, R.E., 1979. Attitude organization and the attitude–behavior relationship. *Journal of personality and social psychology*, 37(6), p.913.

Bailenson, J., Yee, N., Blascovich, J., Beall, A., Lundblad, N. and Jin, M., 2008. The Use of Immersive Virtual Reality in the Learning Sciences: Digital Transformations of Teachers, Students, and Social Context. *Journal of the Learning Sciences*, 17(1), pp.102-141.

Balk, S.A., Bertola, D.B. and Inman, V.W., 2013, June. Simulator sickness questionnaire: twenty years later. In *Driving Assesment Conference* (Vol. 7, No. 2013). University of Iowa.

Bandura, A., 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2), p.191.

Bandura, A., 1986. The explanatory and predictive scope of self-efficacy theory. *Journal of social and clinical psychology*, 4(3), p.359.

Bandura, A., 1997. Self-efficacy: The exercise of control. New York: Freeman.

Bandura, A., 2001. Social cognitive theory: An agentic perspective. *Annual review of psychology*, 52(1), pp.1-26.

Barrett, T. and Hegarty, M., 2016. Effects of interface and spatial ability on manipulation of virtual models in a STEM domain. *Computers in Human Behavior*, 65, pp.220-231.

Bartimote-Aufflick, K., Bridgeman, A., Walker, R., Sharma, M. and Smith, L., 2016. The study, evaluation, and improvement of university student self-efficacy. *Studies in Higher Education*, 41(11), pp.1918-1942.

Bauer, C.F., 2008. Attitude toward chemistry: A semantic differential instrument for assessing curriculum impacts. *Journal of chemical education*, 85(10), p.1440.

Baumeister, J., Ssin, S.Y., ElSayed, N.A., Dorrian, J., Webb, D.P., Walsh, J.A., Simon, T.M., Irlitti, A., Smith, R.T., Kohler, M. and Thomas, B.H., 2017. Cognitive cost of using augmented reality displays. *IEEE transactions on visualization and computer graphics*, 23(11), pp.2378-2388.

Becker, S.A., Cummins, M., Davis, A., Freeman, A., Hall, C.G. and Ananthanarayanan, V., 2017. *NMC horizon report: 2017 higher education edition* (pp. 1-60). The New Media Consortium.

Bellucci, A., Ruiz, A., Díaz, P. and Aedo, I., 2018. Investigating augmented reality support for novice users in circuit prototyping. *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*.

Benbunan-Fich, R. and Hiltz, S.R., 2003. Mediators of the effectiveness of online courses. *IEEE Transactions on Professional communication*, *46*(4), pp.298-312.

Bendis, J.E., 2003. A history and future of stereoscopy in education.

Bennie, S.J., Ranaghan, K.E., Deeks, H., Goldsmith, H.E., O'Connor, M.B., Mulholland, A.J. and Glowacki, D.R., 2019. Teaching enzyme catalysis using interactive molecular dynamics in virtual reality. *Journal of Chemical Education*, *96*(11), pp.2488-2496.

Bernholt, S., Broman, K., Siebert, S. and Parchmann, I., 2018. Digitising Teaching and Learning – Additional Perspectives for Chemistry Education. *Israel Journal of Chemistry*.

Billinghurst, M., Poupyrev, I., Kato, H. and May, R., 2000, July. Mixing realities in shared space: An augmented reality interface for collaborative computing. In 2000 IEEE international conference on multimedia and expo. ICME2000. Proceedings. Latest advances in the fast changing world of multimedia (Cat. No. 00TH8532) (Vol. 3, pp. 1641-1644). IEEE.

Bipp, T., Steinmayr, R. and Spinath, B., 2008. Personality and achievement motivation: Relationship among Big Five domain and facet scales, achievement goals, and intelligence. *Personality and Individual Differences*, 44(7), pp.1454-1464.

Bishop, A.J., 1980. Spatial abilities and mathematics education—A review. *Educational studies in mathematics*, 11(3), pp.257-269.

Bivall, P., Ainsworth, S. and Tibell, L., 2011. Do haptic representations help complex molecular learning?. *Science Education*, 95(4), pp.700-719.

Blessinger, P. and Wankel, C., 2012. Innovative approaches in higher education: an introduction to using immersive interfaces. In *Increasing Student Engagement and Retention Using Immersive Interfaces: Virtual Worlds, Gaming, and Simulation*. Emerald Group Publishing Limited.

Blinder, S.M., 2019. Introduction to the Hartree-Fock method. In *Mathematical Physics in Theoretical Chemistry* (pp. 1-30). Elsevier.

Bloom, B., 1956. *Taxonomy of Educational Objectives*. 2nd ed. Addison-Wesley Longman Ltd.

Bloom, B., Krathwohl, D. and Masia, B., 1973. *Taxonomy of Educational Objectives, the Classification of Educational Goals. Handbook II: Affective Domain*. New York: David McKay Co., Inc.

Bodner, G.M. and Guay, R.B., 1997. The Purdue visualization of rotations test. *The chemical educator*, 2(4), pp.1-17.

Booth, J.L., McGinn, K.M., Young, L.K. and Barbieri, C., 2015. Simple practice doesn't always make perfect: Evidence from the worked example effect. *Policy Insights from the Behavioral and Brain Sciences*, 2(1), pp.24-32.

Borges, M., Symington, A., Coltin, B., Smith, T. and Ventura, R., 2018, October. HTC vive: Analysis and accuracy improvement. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 2610-2615). IEEE.

Brandriet, A.R., Xu, X., Bretz, S.L. and Lewis, J.E., 2011. Diagnosing changes in attitude in first-year college chemistry students with a shortened version of Bauer's semantic differential. *Chemistry Education Research and Practice*, 12(2), pp.271-278.

Bransford, J.D., Brown, A.L. and Cocking, R.R., 2000. *How people learn* (Vol. 11). Washington, DC: National academy press.

Braun, V. and Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), pp.77-101.

Bressler, D.M. and Bodzin, A.M., 2013. A mixed methods assessment of students' flow experiences during a mobile augmented reality science game. *Journal of computer assisted learning*, 29(6), pp.505-517.

Britner, S.L. and Pajares, F., 2001. Self-efficacy beliefs, motivation, race, and gender in middle school science. *Journal of women and Minorities in Science and Engineering*, 7(4).

Bronack, S.C., 2011. The role of immersive media in online education. *The Journal of Continuing Higher Education*, 59(2), pp.113-117.

Brooks, B.J. and Koretsky, M.D., 2011. The influence of group discussion on students’ responses and confidence during peer instruction. *Journal of Chemical Education*, 88(11), pp.1477-1484.

Brown, S.J., White, S., Sharma, B., Wakeling, L., Naiker, M., Chandra, S., Gopalan, R. and Bilimoria, V., 2015. Attitude to the Study of Chemistry and Its Relationship with Achievement in an Introductory Undergraduate Course. *Journal of the Scholarship of Teaching and Learning*, 15(2), pp.33-41.

Bruck, A.D. and Towns, M.H., 2009. Analysis of classroom response system questions via four lenses in a general chemistry course. *Chemistry Education Research and Practice*, 10(4), pp.291-295.

Buchner, J., Buntins, K. and Kerres, M., 2022. The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning*, *38*(1), pp.285-303.

Buchner, J. and Zumbach, J., 2018. Promoting Intrinsic Motivation with a Mobile Augmented Reality Learning Environment. *International Association for Development of the Information Society*.

Bujak, K., Radu, I., Catrambone, R., MacIntyre, B., Zheng, R. and Golubski, G., 2013. A psychological perspective on augmented reality in the mathematics classroom. *Computers & Education*, 68, pp.536-544.

Burrows, A., Holman, J., Lancaster, S., Parsons, A., Overton, T., Pilling, G. and Price, G., 2021. *Chemistry3: Introducing inorganic, organic and physical chemistry*. Oxford University Press.

Cai, S., Liu, C., Wang, T., Liu, E. and Liang, J.C., 2021. Effects of learning physics using Augmented Reality on students’ self-efficacy and conceptions of learning. *British Journal of Educational Technology*, *52*(1), pp.235-251.

Cai, S., Wang, X. and Chiang, F., 2014. A case study of Augmented Reality simulation system application in a chemistry course. *Computers in Human Behavior*, 37, pp.31-40.

Cain, J., 2019. Exploratory implementation of a blended format escape room in a large enrollment pharmacy management class. *Currents in Pharmacy Teaching and Learning*, 11(1), pp.44-50.

Carlisle, D., Tyson, J. and Nieswandt, M., 2015. Fostering spatial skill acquisition by general chemistry students. *Chemistry Education Research and Practice*, 16(3), pp.478-517.

Carnegie, K. and Rhee, T., 2015. Reducing visual discomfort with HMDs using dynamic depth of field. *IEEE computer graphics and applications*, 35(5), pp.34-41.

Carroll, J.B., 1993. *Human cognitive abilities: A survey of factor-analytic studies* (No. 1). Cambridge University Press.

Carson, D., Gilmore, A., Perry, C. and Gronhaug, K., 2001. Qualitative Marketing Research.

Carter, C.S., Larussa, M.A. and Bodner, G.M., 1987. A study of two measures of spatial ability as predictors of success in different levels of general chemistry. *Journal of research in science teaching*, 24(7), pp.645-657.

Caudell, T. and Mizell, D., 1992. Augmented reality: an application of heads-up display technology to manual manufacturing processes. *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*.

Chan, C. and Fok, W., 2009. Evaluating learning experiences in virtual laboratory training through student perceptions: a case study in Electrical and Electronic Engineering at the University of Hong Kong. *Engineering Education*, 4(2), pp.70-75.

Chandler, P. and Sweller, J., 1991. Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, 8(4), pp.293-332.

Chandler, P. and Sweller, J., 1992. The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62(2), pp.233-246.

Chang, S.C. and Hwang, G.J., 2018. Impacts of an augmented reality-based flipped learning guiding approach on students' scientific project performance and perceptions. *Computers & Education*, 125, pp.226-239.

*Chemistry International -- Newsmagazine for IUPAC*, 2005. Nomenclature of Inorganic Chemistry – IUPAC Recommendations 2005. 27(6).

Chen, O., Kalyuga, S. and Sweller, J., 2015. The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology*, 107(3), p.689.

Chen, O., Kalyuga, S. and Sweller, J., 2017. The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review*, 29(2), pp.393-405.

Chen, O., Retnowati, E. and Kalyuga, S., 2020. Element interactivity as a factor influencing the effectiveness of worked example–problem solving and problem solving–worked example sequences. *British Journal of Educational Psychology*, 90, pp.210-223.

Chen, O., Woolcott, G. and Sweller, J., 2017. Using cognitive load theory to structure computer-based learning including MOOCs. *Journal of Computer Assisted Learning*, 33(4), pp.293-305.

Chen, P., Liu, X., Cheng, W. and Huang, R., 2017. A review of using Augmented Reality in Education from 2011 to 2016. *Innovations in smart learning*, pp.13-18.

Chen, Y.-C., Dong, X., Chen, F.-C., & Stoffregen, T. A. (2012). Control of a virtual avatar influences postural activity and motion sickness. *Ecological Psychology*, 24, 279–299. doi:10.1080/10407413.2012.726181

Chen, Y.C., Wang, S.J. and Chiang, Y.L., 2009. Exploring the Effect of Presence in an AR-based Learning Environment. In *13th Global Chinese Conference on Computers in Education, Taipei*.

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

Chou, S.W. and Liu, C.H., 2005. Learning effectiveness in a Web-based virtual learning environment: a learner control perspective. *Journal of computer assisted learning*, *21*(1), pp.65-76.

Clark, R., 1994. Media will never influence learning. *Educational Technology Research and Development*, 42(2), pp.21-29.

Clark, R.C. and Mayer, R.E., 2016. *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. john Wiley & sons.

Clarke, S., Peel, D.J., Arnab, S., Morini, L., Keegan, H. and Wood, O., 2017. EscapED: A framework for creating educational escape rooms and interactive games for higher/further education. *International Journal of Serious Games*, 4(3), pp.73-86.

Cohen, J., 2013. *Statistical power analysis for the behavioral sciences*. Routledge.

Cohen, L., Manion, L. and Morrison, K., 2017. Research Methods in Education.

Coimbra, M.T., Cardoso, T. and Mateus, A., 2015. Augmented reality: an enhancer for higher education students in math's learning?. *Procedia Computer Science*, *67*, pp.332-339.

Coll, R.K., Dalgety, J. and Salter, D., 2002. The development of the chemistry attitudes and experiences questionnaire (CAEQ). *Chemistry Education Research and Practice in Europe*, 3(1), pp.19-32.

Collins, A. and Ferguson, W., 1993. Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational psychologist*, 28(1), pp.25-42.

Cooper, G. and Sweller, J., 1987. Effects of schema acquisition and rule automation on mathematical problem-solving transfer. Journal of educational psychology, 79(4), p.347.

Cortina, J.M., 1993. What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, *78*(1), p.98.

Costa, M.J., 2007. Carbohydeck: A card game to teach the stereochemistry of carbohydrates. *Journal of Chemical Education*, 84(6), p.977.

Cousin, G., 2006. An introduction to threshold concepts. *Planet*, 17(1), pp.4-5.

Cowan, N., 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, 24(1), pp.87-114.

Creswell, J. and Plano Clark, V., 2011. *Designing and conducting mixed methods research*. 2nd ed. Sage Publications.

Creswell, J.W., Plano Clark, V., Gutmann, M, and Hanson, W., 2003. Advanced mixed methods research designs. In A. Tashakkori and C Teddle (Eds.), *Handbook of mixed methods in social and behavioral research,* pp.209-240.

Crippen, K.J. and Brooks, D.W., 2009. Applying cognitive theory to chemistry instruction: the case for worked examples. *Chemistry Education Research and Practice*, 10(1), pp.35-41.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3), pp.297-334.

Crosier, J.K., Cobb, S.V. and Wilson, J.R., 2000. Experimental comparison of virtual reality with traditional teaching methods for teaching radioactivity. *Education and Information Technologies*, *5*(4), pp.329-343.

Crouch, C.H. and Mazur, E., 2001. Peer instruction: Ten years of experience and results. *American journal of physics*, 69(9), pp.970-977.

Dalgarno, B., Hedberg, J. and Harper, B., 2002. The contribution of 3D environments to conceptual understanding.

Dalgarno, B. and Lee, M.J., 2010. What are the learning affordances of 3-D virtual environments?. *British Journal of Educational Technology*, *41*(1), pp.10-32.

Davis, F., 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), p.319.

Davis, F.D., Bagozzi, R.P. and Warshaw, P.R., 1989. User acceptance of computer technology: A comparison of two theoretical models. *Management science*, *35*(8), pp.982-1003.

de Back, T.T., Tinga, A.M., Nguyen, P. and Louwerse, M.M., 2020. Benefits of immersive collaborative learning in CAVE-based virtual reality. *International Journal of Educational Technology in Higher Education*, *17*(1), pp.1-18.

da Silva Júnior, J.N., Sousa Lima, M.A., Xerez Moreira, J.V., Oliveira Alexandre, F.S., de Almeida, D.M., de Oliveira, M.D.C.F. and Melo Leite Junior, A.J., 2017. Stereogame: an interactive computer game that engages students in reviewing stereochemistry concepts.

da Silva Júnior, J.N., Uchoa, D.E.D.A., Sousa Lima, M.A. and Monteiro, A.J., 2019. Stereochemistry game: Creating and playing a fun board game to engage students in reviewing stereochemistry concepts. *Journal of chemical education*, 96(8), pp.1680-1685.

de Souza Silva, T., Marinho, E.C.R., Cabral, G.R.E. and da Gama, K.S., 2017, November. Motivational impact of virtual reality on game-based learning: Comparative study of immersive and non-immersive approaches. In *2017 19th Symposium on Virtual and Augmented Reality* (SVR) (pp. 155-158). IEEE.

Dean, N.L., Ewan, C. and McIndoe, J.S., 2016. Applying hand-held 3D printing technology to the teaching of VSEPR theory.

Deci, E.L. and Ryan, R.M., 1985. The general causality orientations scale: Self-determination in personality. *Journal of research in personality*, 19(2), pp.109-134.

Deci, E.L. and Ryan, R.M., 2008. Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian psychology/Psychologie canadienne*, 49(3), p.182.

Deci, E.L. and Ryan, R.M., 2014. Autonomy and need satisfaction in close relationships: Relationships motivation theory. *Human motivation and interpersonal relationships*, pp.53-73.

Deci, E.L., Schwartz, A.J., Sheinman, L. and Ryan, R.M., 1981. An instrument to assess adults' orientations toward control versus autonomy with children: Reflections on intrinsic motivation and perceived competence. *Journal of educational Psychology*, 73(5), p.642.

Dede, C., 2009. Immersive Interfaces for Engagement and Learning. *Science*, 323(5910), pp.66-69.

Deshpande, A. and Kim, I., 2018. The effects of augmented reality on improving spatial problem solving for object assembly. Advanced Engineering Informatics, 38, pp.760-775.

DeVellis, R., 2006. Classical Test Theory. *Medical Care*, 44(Suppl 3), pp.S50-S59.

Dewey, J., 1910. How we think.

Dewey, J., 1938. *Experience and education*. 1st ed. Kappa Delta Pi.

Dey, A., Billinghurst, M., Lindeman, R.W. and Swan, J.E., 2018. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI*, *5*, p.37.

Di Serio, Á., Ibáñez, M. and Kloos, C., 2013. Impact of an augmented reality system on students' motivation for a visual art course. *Computers & Education*, 68, pp.586-596.

Dietrich, N., 2018. Escape classroom: The leblanc process—An educational "escape game". *Journal of chemical education*, 95(6), pp.996-999.

Dietrich, N., 2019. Chem and roll: A roll and write game to illustrate chemical engineering and the contact process. *Journal of Chemical Education*, 96(6), pp.1194-1198.

Ding, L., Reay, N., Lee, A. and Bao, L., 2011. Exploring the role of conceptual scaffolding in solving synthesis problems. *Physical Review Special Topics-Physics Education Research*, 7(2), p.020109.

Dolin, J., Black, P., Harlen, W. and Tiberghien, A., 2018. Exploring relations between formative and summative assessment. In *Transforming assessment* (pp. 53-80). Springer, Cham.

Domínguez Alfaro, J.L., Gantois, S., Blattgerste, J., De Croon, R., Verbert, K., Pfeiffer, T. and Van Puyvelde, P., 2022. Mobile Augmented Reality Laboratory for Learning Acid–Base Titration.

Donally, J., 2018. Learning transported. *Augmented, virtual, and mixed reality for all classrooms. International Society for Technology in Education,* Portland.

DuBois, J., 2002. When Is Informed Consent Appropriate in Educational Research?: Regulatory and Ethical Issues. *IRB: Ethics and Human Research*, 24(1), p.1.

Duffy, T. and Jonassen, D., 1992. *Constructivism and the Technology of Instruction*. 1st ed. Routledge.

Dunleavy, M. and Dede, C., 2014. Augmented reality teaching and learning. *Handbook of research on educational communications and technology*, pp.735-745.

Dunning, D., Johnson, K., Erlinger, J., & Kruger, J., 2003. Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12, 83–87

Dzulkifli, M.A. and Mustafar, M.F., 2013. The influence of colour on memory performance: A review. *The Malaysian journal of medical sciences*: MJMS, 20(2), p.3.

Edwards, B., Bielawski, K., Prada, R. and Cheok, A., 2019. Haptic virtual reality and immersive learning for enhanced organic chemistry instruction. *Virtual Reality*, 23(4), pp.363-373.

Eilks, I., Moellering, J. and Valanides, N., 2007. Seventh-grade Students' Understanding of Chemical Reactions: Reflections from an Action Research Interview Study. *EURASIA Journal of Mathematics, Science and Technology Education*, 3(4).

Ekstrome, R., French, J., Harman, H. and Dermen, D., 1976. Manual for kit of factor referenced cognitive tests. *Princeton, New Jersey: Educational Testing Service*, *1976*.

Elford, D., Lancaster, S. and Jones, G., 2021. Stereoisomers, Not Stereo Enigmas: A Stereochemistry Escape Activity Incorporating Augmented and Immersive Virtual Reality. *Journal of Chemical Education*, 98(5), pp.1691-1704.

Elford, D., Lancaster, S. and Jones, G., 2022. Exploring the Effect of Augmented Reality on Cognitive Load, Attitude, Spatial Ability, and Stereochemical Perception. *Journal of Science Education and Technology*, 31(3), pp.322-339.

Elford, D., Lancaster, S. and Jones, G., 2022. Fostering Motivation toward Chemistry through Augmented Reality Educational Escape Activities. A Self-Determination Theory Approach. *Journal of Chemical Education*.

Embretson, S. and Reise, S., 2000. *Item Response Theory*. Psychology Press.

Eriksen, K., Nielsen, B.E. and Pittelkow, M., 2020. Visualizing 3D molecular structures using an augmented reality app.

Erlina, E., Cane, C. and Williams, D.P., 2018. Prediction! the VSEPR Game: Using Cards and Molecular Model Building to Actively Enhance Students' Understanding of Molecular Geometry. *Journal of Chemical Education*.

Ertmer, P.A., Ottenbreit-Leftwich, A.T., Sadik, O., Sendurur, E. and Sendurur, P., 2012. Teacher beliefs and technology integration practices: A critical relationship. Computers & education, 59(2), pp.423-435.

Estudante, A. and Dietrich, N., 2020. Using augmented reality to stimulate students and diffuse escape game activities to larger audiences. *Journal of Chemical Education*, 97(5), pp.1368-1374.

Eukel, H. and Morrell, B., 2021. Ensuring educational escape-room success: the process of designing, piloting, evaluating, redesigning, and re-evaluating educational escape rooms. *Simulation & Gaming*, 52(1), pp.18-23.

Ewais, A. and Troyer, O.D., 2019. A usability and acceptance evaluation of the use of augmented reality for learning atoms and molecules reaction by primary school female students in Palestine. *Journal of Educational Computing Research*, 57(7), pp.1643-1670.

Fatemah, A., Rasool, S. and Habib, U., 2020. Interactive 3D visualization of chemical structure diagrams embedded in text to aid spatial learning process of students. *Journal of Chemical Education*, *97*(4), pp.992-1000.

Fencl, H.S. and Scheel, K.R., 2004, September. Pedagogical approaches, contextual variables, and the development of student self-efficacy in undergraduate physics courses. In *AIP Conference Proceedings* (Vol. 720, No. 1, pp. 173-176). American Institute of Physics.

Fernandes, A.S. and Feiner, S.K., 2016, March. Combating VR sickness through subtle dynamic field-of-view modification. In *2016 IEEE symposium on 3D user interfaces (3DUI)* (pp. 201-210). IEEE.

Ferreiro-González, M., Amores-Arrocha, A., Espada-Bellido, E., Aliaño-Gonzalez, M.J., Vázquez-Espinosa, M., González-de-Peredo, A.V., Sancho-Galán, P., Álvarez-Saura, J.Á., Barbero, G.F. and Cejudo-Bastante, C., 2019. Escape classroom: Can you solve a crime using the analytical process?. *Journal of Chemical Education*, 96(2), pp.267-273.

Ferrell, J.B., Campbell, J.P., McCarthy, D.R., McKay, K.T., Hensinger, M., Srinivasan, R., Zhao, X., Wurthmann, A., Li, J. and Schneebeli, S.T., 2019. Chemical exploration with virtual reality in organic teaching laboratories. *Journal of Chemical Education*, *96*(9), pp.1961-1966.

Festinger, L., 1957. *A theory of cognitive dissonance* (Vol. 2). Stanford university press.

Finch, W.H. and French, B.F., 2015. *Latent variable modeling with R*. Routledge.

Fiorella, L., van Gog, T., Hoogerheide, V. and Mayer, R.E., 2017. It’s all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology*, *109*(5), p.653.

Fischer, C.F., 1987. General hartree-fock program. *Computer physics communications*, *43*(3), pp.355-365.

Fischer, M., Fuerst, B., Lee, S.C., Fotouhi, J., Habert, S., Weidert, S., Euler, E., Osgood, G. and Navab, N., 2016. Preclinical usability study of multiple augmented reality concepts for K-wire placement. *International journal of computer assisted radiology and surgery*, *11*(6), pp.1007-1014.

Fombona-Pascual, A., Fombona, J. and Vicente, R., 2022. Augmented Reality, a Review of a Way to Represent and Manipulate 3D Chemical Structures. *Journal of chemical information and modeling*, 62(8), pp.1863-1872.

Fonseca, D. and García-Peñalvo, F.J., 2019. Interactive and collaborative technological ecosystems for improving academic motivation and engagement. *Universal Access in the Information Society*, 18(3), pp.423-430.

Fotaris, P. and Mastoras, T., 2019, October. Escape rooms for learning: A systematic review. In *Proceedings of the European Conference on Games Based Learning* (pp. 235-243).

Foundation, B., 2022. blender.org - Home of the Blender project - Free and Open 3D Creation Software. [online] blender.org. Available at: <https://www.blender.org/> [Accessed 5 September 2022].

Fowler, C., 2015. Virtual reality and learning: Where is the pedagogy?. *British Journal of Educational Technology*, 46(2), pp.412-422.

Fox, R., 2001. Constructivism Examined. *Oxford Review of Education*, 27(1), pp.23-35.

Fraga-Lamas, P., Fernandez-Carames, T., Blanco-Novoa, O. and Vilar-Montesinos, M., 2018. A Review on Industrial Augmented Reality Systems for the Industry 4.0 Shipyard. *IEEE Access*, 6, pp.13358-13375.

Freedman, M.P., 1997. Relationship among laboratory instruction, attitude toward science, and achievement in science knowledge. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *34*(4), pp.343-357.

Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H. and Wenderoth, M.P., 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the national academy of sciences*, *111*(23), pp.8410-8415.

Freund, P.A., Kuhn, J.T. and Holling, H., 2011. Measuring current achievement motivation with the QCM: Short form development and investigation of measurement invariance. *Personality and Individual Differences*, 51(5), pp.629-634.

Frevert, M. and Di Fuccia, D.S., 2019, October. Virtual reality as a means of teaching contemporary chemistry. In *Proceedings of the 2019 The 3rd International Conference on Digital Technology in Education* (pp. 34-38).

Friemert, D., Kaufmann, M., Hartmann, U. and Ellegast, R., 2019, July. First impressions and acceptance of order pickers towards using data glasses at a simulated workstation. In *International Conference on Human-Computer Interaction* (pp. 251-265). Springer, Cham.

Fung, F.M., Choo, W.Y., Ardisara, A., Zimmermann, C.D., Watts, S., Koscielniak, T., Blanc, E., Coumoul, X. and Dumke, R., 2019. Applying a virtual reality platform in environmental chemistry education to conduct a field trip to an overseas site.

Gardner, P.L., 1975. Attitudes to science: A review.

Garris, R., Ahlers, R. and Driskell, J., 2002. Games, Motivation, and Learning: A Research and Practice Model. *Simulation & Gaming*, 33(4), pp.441-467.

Garzón, J. and Acevedo, J., 2019. Meta-analysis of the impact of Augmented Reality on students' learning gains. *Educational Research Review*, 27, pp.244-260.

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

Gavish, N., Gutiérrez, T., Webel, S., Rodríguez, J., Peveri, M., Bockholt, U. and Tecchia, F., 2015. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, 23(6), pp.778-798.

Gaussian. 2022a. Gaussian 16 | Gaussian.com. [online] Available at: <https://gaussian.com/gaussian16/> [Accessed 11 September 2022].

Gaussian.com. 2022b. *GaussView 6* | *Gaussian.com*. [online] Available at: <https://gaussian.com/gaussview6/> [Accessed 18 September 2022].

Gaussian.com. 2022c. *Density Functional (DFT) Methods* | *Gaussian.com*. [online] Available at: <https://gaussian.com/dft/> [Accessed 18 September 2022].

Giang, C., Chevalier, M., Negrini, L., Peleg, R., Bonnet, E., Piatti, A. and Mondada, F., 2018, October. Exploring escape games as a teaching tool in educational robotics. In *International Conference EduRobotics* 2016 (pp. 95-106). Springer, Cham.

Gibson, J.J. and Carmichael, L., 1966. *The senses considered as perceptual systems* (Vol. 2, No. 1, pp. 44-73). Boston: Houghton Mifflin.

Gillespie, R.J., 1963. The valence-shell electron-pair repulsion (VSEPR) theory of directed valency. *Journal of Chemical Education*, *40*(6), p.295.

Ginns, P., 2006. Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and instruction*, 16(6), pp.511-525.

Girvan, C. and Savage, T., 2010. Identifying an appropriate pedagogy for virtual worlds: A Communal Constructivism case study. *Computers & Education*, *55*(1), pp.342-349.

Goff, E.E., Mulvey, K.L., Irvin, M.J. and Hartstone-Rose, A., 2018. Applications of augmented reality in informal science learning sites: a review. *Journal of Science Education and Technology*, *27*(5), pp.433-447.

Golde, M.F., McCreary, C.L. and Koeske, R., 2006. Peer instruction in the general chemistry laboratory: Assessment of student learning. *Journal of Chemical Education*, 83(5), p.804.

Goldhammer, F., Martens, T. and Lüdtke, O., 2017. Conditioning factors of test-taking engagement in PIAAC: An exploratory IRT modelling approach considering person and item characteristics. *Large-Scale Assessments in Education*, 5(1), pp.1-25.

Gómez-Urquiza, J.L., Gómez-Salgado, J., Albendín-García, L., Correa-Rodríguez, M., González-Jiménez, E. and Cañadas-De la Fuente, G.A., 2019. The impact on nursing students' opinions and motivation of using a "Nursing Escape Room" as a teaching game: A descriptive study. *Nurse education today*, 72, pp.73-76.

Google Developers. 2022. *Cloud Anchors allow different users to share AR experiences | ARCore | Google Developers*. [online] Available at: <https://developers.google.com/ar/develop/cloud-anchors#unity-ar-foundation> [Accessed 25 September 2022].

Grandi, J.G., Debarba, H.G., Bemdt, I., Nedel, L. and Maciel, A., 2018, March. Design and assessment of a collaborative 3D interaction technique for handheld augmented reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 49-56). IEEE.

Greer, L. and Heaney, P.J., 2004. Real-time analysis of student comprehension: An assessment of electronic student response technology in an introductory earth science course. *Journal of Geoscience Education*, 52(4), pp.345-351.

Griggs, A., Lazzara, E.H., Palmer, E., Fouquet, S., Leverenz, T., Raushel, A. and Doherty, S., 2019, November. Utilizing Games for Learning: Applications of Game-Based Training and Gamification. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 63, No. 1, pp. 2166-2168). Sage CA: Los Angeles, CA: SAGE Publications.

Griol, D., Molina, J. and Callejas, Z., 2014. An approach to develop intelligent learning environments by means of immersive virtual worlds. *Journal of Ambient Intelligence and Smart Environments*, 6(2), pp.237-255.

Große, C.S., 2015. Fostering Modeling Competencies: Benefits of Worked Examples, Problems to Be Solved, and Fading Procedures. *European Journal of Science and Mathematics Education*, *3*(4), pp.364-375.

Gross, B., Bretschneider-Hagemes, M., Stefan, A. and Rissler, J., 2018, July. Monitors vs. smart glasses: A study on cognitive workload of digital information systems on forklift trucks. In *International conference on digital human modeling and applications in health, safety, ergonomics and risk management* (pp. 569-578). Springer, Cham.

Grove, N. and Bretz, S.L., 2007. CHEMX: An instrument to assess students' cognitive expectations for learning chemistry. *Journal of Chemical Education*, 84(9), p.1524.

Grove, N. and Lowery Bretz, S., 2012. A continuum of learning: from rote memorization to meaningful learning in organic chemistry. *Chem. Educ. Res. Pract.*, 13(3), pp.201-208.

Grushow, A. and Reeves, M., 2019. Using Computational Methods To Teach Chemical Principles: Overview. *ACS Symposium Series*, pp.1-10.

Gutiérrez, A., 1996, July. Visualization in 3-dimensional geometry: In search of a framework. In *Pme Conference* (Vol. 1, pp. 1-3).

Habig, S., 2020. Who can benefit from augmented reality in chemistry? Sex differences in solving stereochemistry problems using augmented reality. *British Journal of Educational Technology*, *51*(3), pp.629-644.

Hadie, S.N. and Yusoff, M.S., 2016. Assessing the validity of the cognitive load scale in a problem-based learning setting. *Journal of Taibah University Medical Sciences*, 11(3), pp.194-202.

Hake, R., 1998. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), pp.64-74.

Halcrow, M.A., 2013. Jahn–Teller distortions in transition metal compounds, and their importance in functional molecular and inorganic materials. *Chemical Society Reviews*, *42*(4), pp.1784-1795.

Halpern, D.F. and Hakel, M.D., 2003. Applying the science of learning to the university and beyond: Teaching for long-term retention and transfer. *Change: The Magazine of Higher Learning*, *35*(4), pp.36-41.

Hamari, J., Shernoff, D.J., Rowe, E., Coller, B., Asbell-Clarke, J. and Edwards, T., 2016. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in human behavior*, 54, pp.170-179.

Hambleton, R.K., Swaminathan, H. and Rogers, H.J., 1991. *Fundamentals of item response theory* (Vol. 2). Sage.

Harackiewicz, J.M., Smith, J.L. and Priniski, S.J., 2016. Interest matters: The importance of promoting interest in education. *Policy insights from the behavioral and brain sciences*, *3*(2), pp.220-227.

Hardre, P.L. and Reeve, J., 2003. A motivational model of rural students' intentions to persist in, versus drop out of, high school. *Journal of educational psychology*, 95(2), p.347.

Harle, M. and Towns, M., 2011. A review of spatial ability literature, its connection to chemistry, and implications for instruction. *Journal of Chemical Education*, 88(3), pp.351-360.

Hauptman, H., 2010. Enhancement of spatial thinking with Virtual Spaces 1.0. *Computers & Education*, 54(1), pp.123-135.

Heckhausen, H., 1977. Achievement motivation and its constructs: A cognitive model. *Motivation and emotion*, 1(4), pp.283-329.

Heckman, J.J., 1990. Selection bias and self-selection. In *Econometrics* (pp. 201-224). Palgrave Macmillan, London.

Heilig, M., 1955. The cinema of the future.

Heilig, M., 1960. *Stereoscopic-television apparatus for individual use*. US2955156A.

Helgaker, T. and Taylor, P.R., 1995. Gaussian basis sets and molecular integrals. *Modern Electronic Structure Theory: Part II*, pp.725-856.

Hestenes, D., Wells, M. and Swackhamer, G., 1992. Force concept inventory. *The physics teacher*, 30(3), pp.141-158.

Hew, K. and Cheung, W., 2008. Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British Journal of Educational Technology*, 41(1), pp.33-55.

Hick, W.E., 1952. On the rate of gain of information. *Quarterly Journal of experimental psychology*, 4(1), pp.11-26.

Hilbert, T.S., Renkl, A., Kessler, S. and Reiss, K., 2008. Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction*, 18(1), pp.54-65.

Hinds, P.J., 1999. The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of experimental psychology*: applied, 5(2), p.205.

Hoe, Z., Lee, I., Chen, C. and Chang, K., 2017. Using an augmented reality-based training system to promote spatial visualization ability for the elderly. *Universal Access in the Information Society*.

Hohenberg, P. and Kohn, W., 1964. Inhomogeneous electron gas. *Physical review*, *136*(3B), p.B864.

Hsu, C.Y., Kalyuga, S. and Sweller, J., 2015. When should guidance be presented in physics instruction?. *Archives of Scientific Psychology*, 3(1), p.37.

Hu, Y., Gallagher, T., Wouters, P., van der Schaaf, M. and Kester, L., 2022. Game-based learning has good chemistry with chemistry education: A three-level meta-analysis. *Journal of Research in Science Teaching*.

Huang, Y., Backman, S., Backman, K., McGuire, F. and Moore, D., 2018. An investigation of motivation and experience in virtual learning environments: a self-determination theory. *Education and Information Technologies*, 24(1), pp.591-611.

Huang, Y., Du, C., Xue, Z. and Chen, X., 2021. What Makes Multi-modal Learning Better than Single (Provably). *35th Conference on Neural Information Processing Systems (NeurIPS 2021).*

Huang, W., Roscoe, R.D., Johnson-Glenberg, M.C. and Craig, S.D., 2021. Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion. *Journal of Computer Assisted Learning*, *37*(3), pp.745-758.

Hudson, L. and Ozanne, J., 1988. Alternative Ways of Seeking Knowledge in Consumer Research. *Journal of Consumer Research*, 14(4), p.508.

Humphrey, K., 2017. The application of a serious, non-digital escape game learning experience in higher education. *Sport & Exercise Psychology Review*, *13*(2), pp.48-54.

Hurley, Z., 2022. Thinking with semiotic-dialogism: Re-orientating augmented reality and visual literacy. *Studies in Technology Enhanced Learning*.

Huk, T., 2006. Who benefits from learning with 3D models? The case of spatial ability. *Journal of computer assisted learning*, 22(6), pp.392-404.

Hwang, G.J. and Wu, P.H., 2014. Applications, impacts and trends of mobile technology-enhanced learning: a review of 2008–2012 publications in selected SSCI journals. *International Journal of Mobile Learning and Organisation*, 8(2), pp.83-95.

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

IGI Global, 2022. *What is Interactive Technology*. [online] Igi-global.com. Available at: <https://www.igi-global.com/dictionary/parental-mediation-of-adolescent-technology-use/41845> [Accessed 3 September 2022].

Imai, T., Qui, Z., Behara, S., Tachi, S., Aoyama, T., Johnson, A. and Leigh, J., 2000, July. Overcoming time-zone differences and time management problem with tele-immersion. In *the proceedings of INET* (pp. 18-21).

Imran, E., Adanir, N. and Khurshid, Z., 2021. Significance of Haptic and Virtual Reality Simulation (VRS) in the Dental Education: A Review of Literature. *Applied Sciences*, 11(21), p.10196.

Inkpen, K.M., 1997. *Adapting the human-computer interface to support collaborative learning environments for children*. University of British Columbia.

Irwansyah, F.S., Yusuf, Y.M., Farida, I. and Ramdhani, M.A., 2018. Augmented reality (AR) technology on the android operating system in chemistry learning. In *IOP conference series: Materials science and engineering* (Vol. 288, No. 1, p. 012068). IOP Publishing.

Irwansyah, F.S., Nur Asyiah, E., Maylawati, D.S.A., Farida, I. and Ramdhani, M.A., 2020. The development of augmented reality applications for chemistry learning. In *Augmented reality in education* (pp. 159-183). Springer, Cham.

Irwin, J.R. and McClelland, G.H., 2003. Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, 40(3), pp.366-371.

Izatt, E., Scholberg, K. and Kopper, R., 2014. Neutrino-KAVE: An immersive visualization and fitting tool for neutrino physics education. *2014 IEEE Virtual Reality (VR)*.

Jahn, H.A. and Teller, E., 1937. Stability of polyatomic molecules in degenerate electronic states-I—Orbital degeneracy. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, 161(905), pp.220-235.

Janonis, A., Kiudys, E., Girdžiūna, M., Blažauskas, T., Paulauskas, L. and Andrejevas, A., 2020, October. Escape the lab: Chemical experiments in virtual reality. In *International Conference on Information and Software Technologies* (pp. 273-282). Springer, Cham.

Järveläinen, J. and Paavilainen-Mäntymäki, E., 2019. Escape room as game-based learning process: causation-effectuation perspective.

Janssen, J., Kirschner, F., Erkens, G., Kirschner, P.A. and Paas, F., 2010. Making the black box of collaborative learning transparent: Combining process-oriented and cognitive load approaches. *Educational psychology review*, 22(2), pp.139-154.

Jesionkowska, J., Wild, F. and Deval, Y., 2020. Active Learning Augmented Reality for STEAM Education—A Case Study. *Education Sciences*, 10(8), p.198.

Johnson, L., Levine, A., Smith, R. and Stone, S., 2010. The 2010 Horizon Report.

Johnstone, A., 1982. Macro- and micro-chemistry. *School Science Review*, 64, pp.377-379.

Johnstone, A., 1991. Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, 7(2), pp.75-83.

Jonassen, D., 2009. Reconciling a human cognitive architecture. In *Constructivist instruction* (pp. 25-45). Routledge.

Jones, K., 1998. Theoretical frameworks for the learning of geometrical reasoning. *Proceedings of the British Society for Research into Learning Mathematics*, *18*(1-2), pp.29-34.

Jones, R.M. and Fleischman, E.S., 2001. CASCADE explains and informs the utility of fading examples to problems. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 459-464).

Jung, J., Shin, Y. and Zumbach, J., 2021. The effects of pre-training types on cognitive load, collaborative knowledge construction and deep learning in a computer-supported collaborative learning environment. Interactive Learning Environments, 29(7), pp.1163-1175.

Kahveci, A., 2015. Assessing high school students' attitudes toward chemistry with a shortened semantic differential. *Chemistry Education Research and Practice*, 16(2), pp.283-292.

Kalyuga, S., 2007. Expertise reversal effect and its implications for learner-tailored instruction. *Educational psychology review*, 19(4), pp.509-539.

Kalyuga, S., 2009. The expertise reversal effect. In *Managing cognitive load in adaptive multimedia learning* (pp. 58-80). IGI Global.

Kalyuga, S. and Hanham, J., 2011. Instructing in generalized knowledge structures to develop flexible problem solving skills. *Computers in Human Behavior*, 27(1), pp.63-68.

Kalyuga, S. and Renkl, A., 2010. Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science*, 38(3), pp.209-215.

Kalyuga, S., Chandler, P., Tuovinen, J. and Sweller, J., 2001. When problem solving is superior to studying worked examples. *Journal of educational psychology*, 93(3), p.579.

Kardong-Edgren, S., Farra, S., Alinier, G. and Young, H., 2019. A Call to Unify Definitions of Virtual Reality. *Clinical Simulation in Nursing*, 31, pp.28-34.

Karami, H., 2012. An introduction to differential item functioning. *The International Journal of Educational and Psychological Assessment*.

Kato, H. and Billinghurst, M., 1999, October. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)* (pp. 85-94). IEEE.

Kaufmann, H., 2003. Collaborative augmented reality in education. *Institute of Software Technology and Interactive Systems, Vienna University of Technology*, pp.2-4.

Keehner, M. and Khooshabeh, P., 2005. Computerized Representations of 3D Structure: How Spatial Comprehension and Patterns of Interactivity Differ among Learners. In *AAAI Spring Symposium: Reasoning with Mental and External Diagrams: Computational Modeling and Spatial Assistance* (pp. 12-17).

Keehner, M., Montello, D.R., Hegarty, M. and Cohen, C., 2004. Effects of interactivity and spatial ability on the comprehension of spatial relations in a 3D computer visualization. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 26, No. 26).

Keller, S., Rumann, S. and Habig, S., 2021. Cognitive Load Implications for Augmented Reality Supported Chemistry Learning. *Information*, 12(3), p.96.

Kerawalla, L., Luckin, R., Seljeflot, S. and Woolard, A., 2006. “Making it real”: exploring the potential of augmented reality for teaching primary school science. *Virtual reality*, 10(3), pp.163-174.

Kester, L., Kirschner, P.A. and van Merriënboer, J.J., 2006. Just-in-time information presentation: Improving learning a troubleshooting skill. *Contemporary Educational Psychology*, 31(2), pp.167-185.

Kim, H.K., Park, J., Choi, Y. and Choe, M., 2018. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied ergonomics*, 69, pp.66-73.

Kirschner, F., Paas, F. and Kirschner, P.A., 2009. A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational psychology review*, 21(1), pp.31-42.

Klopfer, E., Squire, K. and Jenkins, H., 2002, August. Environmental detectives: PDAs as a window into a virtual simulated world. In *Proceedings. IEEE international workshop on wireless and mobile technologies in education* (pp. 95-98). IEEE.

Knight, J.K. and Wood, W.B., 2005. Teaching more by lecturing less. *Cell biology education*, 4(4), pp.298-310.

Koch, M., von Luck, K., Schwarzer, J. and Draheim, S., 2018. The Novelty Effect in Large Display Deployments - Experiences and Lessons-Learned for Evaluating Prototypes. *Reports of the European Society for Socially Embedded Technologies*, 2(1).

Kodiyah, J., Irwansyah, F.S. and Windayani, N., 2020, March. Application of augmented reality (AR) media on conformation of alkanes and cycloalkanes concepts to improve student’s spatial ability. In *Journal of Physics: Conference Series* (Vol. 1521, No. 4, p. 042093). IOP Publishing.

Koedinger, K.R. and Aleven, V., 2007. Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), pp.239-264.

Kolasinski, E.M., 1995. *Simulator sickness in virtual environments* (Vol. 1027). US Army Research Institute for the Behavioral and Social Sciences.

Kolb, D.A., 1965. Achievement motivation training for underachieving high-school boys. *Journal of Personality and Social Psychology*, 2(6), p.783.

Kolb, D.A., 2014. *Experiential learning: Experience as the source of learning and development*. FT press.

Koslucher, F., Haaland, E., Malsch, A., Webeler, J. and Stoffregen, T.A., 2015. Sex differences in the incidence of motion sickness induced by linear visual oscillation. *Aerospace medicine and human performance*, 86(9), pp.787-793.

Kozma, R., 1994. Will media influence learning? Reframing the debate. *Educational Technology Research and Development*, 42(2), pp.7-19.

Krajcik, J. S. and Blumenfeld, P. C., 2005. “Project-Based Learning,” in Sawyer, R. K. (ed.) *The Cambridge Handbook of the Learning Sciences*. Cambridge: Cambridge University Press (Cambridge Handbooks in Psychology), pp. 317–334. doi: 10.1017/CBO9780511816833.020.

Krippendorff, K., 2018. *Content analysis: An Introduction to Its Methodology*. 4th ed. SAGE Publications.

Krueger, M.W., Gionfriddo, T. and Hinrichsen, K., 1985, April. VIDEOPLACE—an artificial reality. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 35-40).

Kuhn, T., 1962. *The Structure of Scientific Revolutions*. 2nd ed. Chicago: The University of Chicago Press.

Kuncel, N.R. and Hezlett, S.A., 2010. Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science*, 19(6), pp.339-345.

Kusuma, I.A. and Retnowati, E., 2021, January. Faded-Example Effects in Individual or Group Work Settings. In *2021 10th International Conference on Educational and Information Technology (ICEIT)* (pp. 204-208). IEEE.

Laal, M. and Ghodsi, S., 2012. Benefits of collaborative learning. *Procedia - Social and Behavioral Sciences*, 31, pp.486-490.

Lai, A.F., Chen, C.H. and Lee, G.Y., 2019. An augmented reality-based learning approach to enhancing students’ science reading performances from the perspective of the cognitive load theory. *British Journal of Educational Technology*, 50(1), pp.232-247.

Lampen, E., Teuber, J., Gaisbauer, F., Bär, T., Pfeiffer, T. and Wachsmuth, S., 2019. Combining simulation and augmented reality methods for enhanced worker assistance in manual assembly. *Procedia Cirp*, *81*, pp.588-593.

Lancaster, S., Cook, D.F. and Massingberd-Mundy, W.J., 2019. Peer instruction as a flexible, scalable, active learning approach in higher education. *Teaching Chemistry in Higher Education: A Festschrift in Honour of Professor Tina Overton*, p.89.

Lasry, N., Charles, E. and Whittaker, C., 2016. Effective variations of peer instruction: The effects of peer discussions, committing to an answer, and reaching a consensus. *American Journal of Physics*, *84*(8), pp.639-645.

Lazowski, R.A. and Hulleman, C.S., 2016. Motivation interventions in education: A meta-analytic review. *Review of Educational research*, 86(2), pp.602-640.

Lee, K., 2012. Augmented Reality in Education and Training. *TechTrends*, 56(2), pp.13-21.

Lee, E.A.L., Wong, K.W. and Fung, C.C., 2010. How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education*, *55*(4), pp.1424-1442.

Leontyev, A., 2015. Development of a stereochemistry concept inventory. University of Northern Colorado.

Leppink, J., 2017. Cognitive load theory: Practical implications and an important challenge. *Journal of Taibah University Medical Sciences*, *12*(5), pp.385-391.

Leppink, J., Paas, F., Van der Vleuten, C.P., Van Gog, T. and Van Merriënboer, J.J., 2013. Development of an instrument for measuring different types of cognitive load. *Behavior research methods*, 45(4), pp.1058-1072.

Leppink, J., Paas, F., Van Gog, T., van Der Vleuten, C.P. and Van Merrienboer, J.J., 2014. Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and instruction*, 30, pp.32-42.

Li, M.C. and Tsai, C.C., 2013. Game-based learning in science education: A review of relevant research. *Journal of Science Education and Technology*, 22(6), pp.877-898.

Library.vuforia.com. 2022. *Image Targets | VuforiaLibrary*. [online] Available at: <https://library.vuforia.com/objects/image-targets> [Accessed 30 August 2022].

Lim, C., Ab Jalil, H., Ma'rof, A. and Saad, W., 2020. Peer learning, self-regulated learning and academic achievement in blended learning courses: A structural equation modeling approach. *International Journal of Emerging Technologies in Learning* (IJET), 15(3), pp.110-125.

Lim, K., Selamat, A., Alias, R., Krejcar, O. and Fujita, H., 2019. Usability Measures in Mobile-Based Augmented Reality Learning Applications: A Systematic Review. *Applied Sciences*, 9(13), p.2718.

Limniou, M., Roberts, D. and Papadopoulos, N., 2008. Full immersive virtual environment CAVETM in chemistry education. *Computers & Education*, *51*(2), pp.584-593.

Linnenbrink-Garcia, L., Durik, A.M., Conley, A.M., Barron, K.E., Tauer, J.M., Karabenick, S.A. and Harackiewicz, J.M., 2010. Measuring situational interest in academic domains. *Educational and psychological measurement*, 70(4), pp.647-671.

Lin, T.J., Duh, H.B.L., Li, N., Wang, H.Y. and Tsai, C.C., 2013. An investigation of learners' collaborative knowledge construction performances and behavior patterns in an augmented reality simulation system. Computers & Education, 68, pp.314-321.

Lin, J.W., Duh, H.B.L., Parker, D.E., Abi-Rached, H. and Furness, T.A., 2002, March. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proceedings ieee virtual reality 2002* (pp. 164-171). IEEE.

Liu, G., Zhang, S., Zhang, J., Lee, C., Wang, Y. and Brownell, M., 2013. Autonomous motivation and Chinese adolescents' creative thinking: The moderating role of parental involvement. *Creativity research journal*, 25(4), pp.446-456.

Lohman, D.F., 1988. Spatial abilities as traits, processes, and knowledge.

Lohman, D.F. and Kyllonen, P.C., 1983. Individual differences in solution strategy on spatial tasks. *Individual differences in cognition*, 1, pp.105-135.

Lombard, M. and Ditton, T., 2006. At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, 3(2).

López-Pernas, S., Gordillo, A. and Barra, E., 2021. Technology-enhanced educational escape rooms: A road map. *IT Professional*, 23(2), pp.26-32.

López-Pernas, S., Gordillo, A., Barra, E. and Quemada, J., 2019. Examining the use of an educational escape room for teaching programming in a higher education setting. IEEE Access, 7, pp.31723-31737.

Low, R.C., Jin, P.C. and Sweller, J.C., 2011. *Cognitive load theory, attentional processes and optimized learning outcomes in a digital environment*. Cambridge University Press.

Lu, A., Wong, C.S., Cheung, R.Y. and Im, T.S., 2021, April. Supporting flipped and gamified learning with augmented reality in higher education. In *Frontiers in Education* (Vol. 6, p. 623745). Frontiers Media SA.

Lukosch, S., Billinghurst, M., Alem, L. and Kiyokawa, K., 2015. Collaboration in augmented reality. *Computer Supported Cooperative Work (CSCW)*, *24*(6), pp.515-525.

Lupi, J., Martino, M., Salvadori, A., Rampino, S., Mancini, G. and Barone, V., 2019, August. Virtual reality tools for advanced modeling. In *AIP Conference Proceedings* (Vol. 2145, No. 1, p. 020001). AIP Publishing LLC.

Luxford, C. and Bretz, S., 2014. Development of the Bonding Representations Inventory To Identify Student Misconceptions about Covalent and Ionic Bonding Representations. *Journal of Chemical Education*, 91(3), pp.312-320.

Magis, D., Béland, S., Tuerlinckx, F. and De Boeck, P., 2010. A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior research methods*, 42(3), pp.847-862.

Magson, S.; Macpherson, S. 2022. *All escape games and best escape rooms in the United Kingdom on worldofescapes.co.uk*. [online] Available at: <https://worldofescapes.co.uk/> [Accessed 14 September 2022].

Mahaffy, P., 2006. Moving Chemistry Education into 3D: A Tetrahedral Metaphor for Understanding Chemistry. Union Carbide Award for Chemical Education. *Journal of Chemical Education*, 83(1), p.49.

Maier, P. and Klinker, G., 2013, September. Augmented chemical reactions: An augmented reality tool to support chemistry teaching. In *2013 2nd Experiment@ International Conference (exp. at'13)* (pp. 164-165). IEEE.

Makransky, G. and Lilleholt, L., 2018. A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, 66(5), pp.1141-1164.

Malone, T.W., 1981. Toward a theory of intrinsically motivating instruction. *Cognitive science*, 5(4), pp.333-369.

Mao, C.C., Chen, C.H. and Sun, C.C., 2017. Impact of an augmented reality system on learning for army military decision-making process (MDMP) course. In *Advances in Ergonomics Modeling, Usability & Special Populations* (pp. 663-671). Springer, Cham.

Maor, D. and Fraser, B.J., 2005. An online questionnaire for evaluating students' and teachers' perceptions of constructivist multimedia learning environments. *Research in Science Education*, *35*(2), pp.221-244.

Marks, R.B., Sibley, S.D. and Arbaugh, J.B., 2005. A structural equation model of predictors for effective online learning. *Journal of management education*, *29*(4), pp.531-563.

Martens, R., Bastiaens, T. and Kirschner, P.A., 2007. New learning design in distance education: The impact on student perception and motivation. *Distance education*, *28*(1), pp.81-93.

Martin, C.B., Vandehoef, C. and Cook, A., 2015. The use of molecular modeling as “pseudoexperimental” data for teaching VSEPR as a hands-on general chemistry activity. *Journal of Chemical Education*, 92(8), pp.1364-1368.

Martin, D., Malpica, S., Gutierrez, D., Masia, B. and Serrano, A., 2022. Multimodality in VR: A survey. *ACM Computing Surveys*,.

Martín-Gutiérrez, J., Mora, C.E., Añorbe-Díaz, B. and González-Marrero, A., 2017. Virtual technologies trends in education. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(2), pp.469-486.

Marx, J. and Cummings, K., 2007. Normalized change. *American Journal of Physics*, 75(1), pp.87-91.

Mascolo, M.F., Fischer, K. W., 2005. *Constructivist theories. Cambridge Encyclopedia of Child Development*. Cambridge University Press, pp.49-63.

Mayer, R., 2002. Cognitive Theory and the Design of Multimedia Instruction: An Example of the Two-Way Street Between Cognition and Instruction. *New Directions for Teaching and Learning*, 2002(89), pp.55-71.

Mayer, R.E., 2005. Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, 41, pp.31-48.

Mayer, R.E., 2019. Computer games in education. *Annual review of psychology*, *70*, pp.531-549.

Mayer, R.E. and Fiorella, L., 2014. 12 principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In *The Cambridge handbook of multimedia learning* (Vol. 279). New York, NY: Cambridge University Press.

Mayer, R. E., Mathias, A., & Wetzell, K., 2002. Fostering understanding of multimedia messages through pretraining: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied*, 8, pp.147–154.

Mayer, R.E., Mautone, P. and Prothero, W., 2002. Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology*, *94*(1), p.171.

Mayer, R.E. and Moreno, R., 1998. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of educational psychology*, 90(2), p.312.

Mayer, R.E. and Pilegard, C., 2005. Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. The Cambridge handbook of multimedia learning, pp.169-182.

Mayer, R.E. and Sims, V.K., 1994. For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of educational psychology*, 86(3), p.389.

Mazur, E., 1997, March. Peer instruction: Getting students to think in class. In AIP Conference Proceedings (Vol. 399, No. 1, pp. 981-988). *American Institute of Physics*.

Mazzuco, A., Krassmann, A., Reategui, E. and Gomes, R., 2022. A systematic review of augmented reality in chemistry education. *Review of Education*, 10(1).

McAuley, E., Duncan, T. and Tammen, V.V., 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport*, 60(1), pp.48-58.

McClelland, D., 1961. *The achieving society*.

McConnell, D.A., Steer, D.N. and Owens, K.D., 2003. Assessment and active learning strategies for introductory geology courses. *Journal of Geoscience Education*, 51(2), pp.205-216.

McConnell, D.A., Steer, D.N., Owens, K.D., Knott, J.R., Van Horn, S., Borowski, W., Dick, J., Foos, A., Malone, M., McGrew, H. and Greer, L., 2006. Using conceptests to assess and improve student conceptual understanding in introductory geoscience courses. *Journal of Geoscience Education*, 54(1), pp.61-68.

McConville, K.M.V. and Milosevic, M., 2014. Active video game head movement inputs. *Personal and ubiquitous computing*, 18(1), pp.253-257.

McGee, M.G., 1979. Human spatial abilities: psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological bulletin*, 86(5), p.889.

McGrath, C.H., Guerin, B., Harte, E., Frearson, M. and Manville, C., 2015. Learning gain in higher education. Santa Monica, CA: RAND Corporation.

McKenney, S. and Reeves, T.C., 2014. Educational design research. In Handbook of research on educational communications and technology (pp. 131-140). Springer, New York, NY.

McKernan, B., Martey, R.M., Stromer-Galley, J., Kenski, K., Clegg, B.A., Folkestad, J.E., Rhodes, M.G., Shaw, A., Saulnier, E.T. and Strzalkowski, T., 2015. We don't need no stinkin'badges: The impact of reward features and feeling rewarded in educational games. *Computers in Human Behavior*, 45, pp.299-306.

Merchant, Z., Goetz, E.T., Keeney-Kennicutt, W., Cifuentes, L., Kwok, O.M. and Davis, T.J., 2013. Exploring 3-D virtual reality technology for spatial ability and chemistry achievement. *Journal of Computer Assisted Learning*, 29(6), pp.579-590.

Messick, S., 1994. The interplay of evidence and consequences in the validation of performance assessments. *Educational researcher*, 23(2), pp.13-23.

Mestre, D. and Vercher, J., 2011. Immersion and presence.

Meyer, O.A., Omdahl, M.K. and Makransky, G., 2019. Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, 140, p.103603.

Michael, W.B., Guilford, J.P., Fruchter, B. and Zimmerman, W.S., 1957. The description of spatial-visualization abilities. *Educational and psychological measurement*, *17*(2), pp.185-199.

Microsoft.com. 2022. Video Conferencing, Meetings, Calling | Microsoft Teams. [online] Available at: <https://www.microsoft.com/en-gb/microsoft-teams/group-chat-software> [Accessed 5 September 2022].

Milgram, P., Takemura, H., Utsumi, A. and Kishino, F., 1994. Augmented reality: a class of displays on the reality-virtuality continuum. *SPIE Proceedings*.

Miller, K., Schell, J., Ho, A., Lukoff, B. and Mazur, E., 2015. Response switching and self-efficacy in Peer Instruction classrooms. *Physical Review Special Topics-Physics Education Research*, 11(1), p.010104.

Miller, R.L., Santana-Vega, E. and Terrell, M.S., 2006. Can good questions and peer discussion improve calculus instruction?. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *16*(3), pp.193-203.

Mishra, P., Pandey, C.M., Singh, U., Gupta, A., Sahu, C. and Keshri, A., 2019. Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, *22*(1), p.67.

Moore, P., 1995. Learning and teaching in virtual worlds: Implications of virtual reality for education. *Australasian Journal of Educational Technology*, *11*(2).

Moreno, R. and Mayer, R., 2007. Interactive multimodal learning environments. *Educational psychology review*, 19(3), pp.309-326.

Moreno, R.E. and Park, B., 2010. Cognitive load theory: Historical development and relation to other theories.

Moritz, S.E., Feltz, D.L., Fahrbach, K.R. and Mack, D.E., 2000. The relation of self-efficacy measures to sport performance: A meta-analytic review. *Research quarterly for exercise and sport*, 71(3), pp.280-294.

Mulford, D. and Robinson, W., 2002. An Inventory for Alternate Conceptions among First-Semester General Chemistry Students. *Journal of Chemical Education*, 79(6), p.739.

Munafo, J., Diedrick, M. and Stoffregen, T.A., 2017. The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. *Experimental brain research*, 235(3), pp.889-901.

Munby, H., 1983. Thirty studies involving the “Scientific Attitude Inventory”: What confidence can we have in this instrument?. *Journal of research in science teaching*, 20(2), pp.141-162.

Mwangi, W. and Sweller, J., 1998. Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and instruction*, 16(2), pp.173-199.

Naese, J.A., McAteer, D., Hughes, K.D., Kelbon, C., Mugweru, A. and Grinias, J.P., 2019. Use of augmented reality in the instruction of analytical instrumentation design.

Nanome.ai. 2022. Nanome: Virtual Reality for Drug Design and Molecular Visualization. [online] Available at: <https://nanome.ai/> [Accessed 5 September 2022].

Nazar, M., Aisyi, R., Rahmayani, R.F.I., Hanum, L., Rusman, R., Puspita, K. and Hidayat, M., 2020, February. Development of augmented reality application for learning the concept of molecular geometry. In *Journal of Physics: Conference Series* (Vol. 1460, No. 1, p. 012083). IOP Publishing.

Neumann, K.L., Alvarado-Albertorio, F. and Ramírez-Salgado, A., 2020. Online approaches for implementing a digital escape room with preservice teachers. *Journal of Technology and Teacher Education*, 28(2), pp.415-424.

Newbury, P., 2014. *Preparing to Teach 3: Sample Peer Instruction Questions*.

Nicholson, S., 2015. A recipe for meaningful gamification. In *Gamification in education and business* (pp. 1-20). Springer, Cham.

Nicholson, S. and Cable, L., 2021. Unlocking the potential of puzzle-based learning.

Nicol, D.J. and Boyle, J.T., 2003. Peer instruction versus class-wide discussion in large classes: A comparison of two interaction methods in the wired classroom. *Studies in higher education*, 28(4), pp.457-473.

Nielsen, K.L., Hansen, G. and Stav, J.B., 2016. How the initial thinking period affects student argumentation during peer instruction: students' experiences versus observations. *Studies in higher education*, *41*(1), pp.124-138.

Nielsen, K.L., Hansen-Nygård, G. and Stav, J.B., 2012. Investigating peer instruction: How the initial voting session affects students' experiences of group discussion. *International Scholarly Research Notices*, *2012*.

Nitta, H., 2010. Mathematical theory of peer-instruction dynamics. *Physical Review Special Topics-Physics Education Research*, 6(2), p.020105.

Nitta, H., Matsuura, S. and Kudo, T., 2014. Implementation and analysis of peer-instruction in introductory physics lectures. *Journal of Science Education in Japan*, 38, pp.12-19.

Norman, K.L., 1994. Spatial visualization—a gateway to computer-based technology. *Journal of Special Education Technology*, 12(3), pp.195-206.

Norrby, M., Grebner, C., Eriksson, J. and Bostrom, J., 2015. Molecular rift: virtual reality for drug designers. *Journal of chemical information and modeling*, *55*(11), pp.2475-2484.

Núñez, M., Quirós, R., Núñez, I., Carda, J.B., Camahort, E. and Mauri, J.L., 2008, July. Collaborative augmented reality for inorganic chemistry education. In *WSEAS international conference. Proceedings. Mathematics and computers in science and engineering* (Vol. 5, pp. 271-277). WSEAS.

O'Connor, M.B., Bennie, S.J., Deeks, H.M., Jamieson-Binnie, A., Jones, A.J., Shannon, R.J., Walters, R., Mitchell, T.J., Mulholland, A.J. and Glowacki, D.R., 2019. Interactive molecular dynamics in virtual reality from quantum chemistry to drug binding: An open-source multi-person framework. *The Journal of chemical physics*, *150*(22), p.220901.

O'Connor, M., Deeks, H., Dawn, E., Metatla, O., Roudaut, A., Sutton, M., Thomas, L., Glowacki, B., Sage, R., Tew, P., Wonnacott, M., Bates, P., Mulholland,

A. and Glowacki, D., 2018. Sampling molecular conformations and dynamics in a multiuser virtual reality framework. *Science Advances*, 4(6).

O'Shea, H. and Moran, A., 2017. Does motor simulation theory explain the cognitive mechanisms underlying motor imagery? A critical review. *Frontiers in Human Neuroscience*, *11*, p.72.

Oestreich, J.H., Hunt, B. and Cain, J., 2021. Grant deadline: An escape room to simulate grant submissions. *Currents in Pharmacy Teaching and Learning*, 13(7), pp.848-854.

Ohyama, S., Nishiike, S., Watanabe, H., Matsuoka, K., Akizuki, H., Takeda, N. and Harada, T., 2007. Autonomic responses during motion sickness induced by virtual reality. *Auris Nasus Larynx*, 34(3), pp.303-306.

Olim, S.C. and Nisi, V., 2020, November. Augmented reality towards facilitating abstract concepts learning. In *International Conference on Entertainment Computing* (pp. 188-204). Springer, Cham.

Olsson, P., Nysjo, F., Seipel, S. and Carlbom, I., 2012. Physically co-located haptic interaction with 3D displays. *2012 IEEE Haptics Symposium (HAPTICS)*.

Olympiou, G. and Zacharia, Z.C., 2012. Blending physical and virtual manipulatives: An effort to improve students' conceptual understanding through science laboratory experimentation. *Science Education*, 96(1), pp.21-47.

Openbabel.org. 2022. Open Babel. [online] Available at: <http://openbabel.org/wiki/Main_Page> [Accessed 5 September 2022].

Orlando, M. and Thissen, D., 2003. Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, *27*(4), pp.289-298.

Otis, N., Grouzet, F.M. and Pelletier, L.G., 2005. Latent motivational change in an academic setting: a 3-year longitudinal study. *Journal of educational psychology*, 97(2), p.170.

Oxford, R.L., 1997. Cooperative learning, collaborative learning, and interaction: Three communicative strands in the language classroom. *The modern language journal*, 81(4), pp.443-456.

Ozdemir, M., Sahin, C., Arcagok, S. and Demir, M., 2018. The Effect of Augmented Reality Applications in the Learning Process: A Meta-Analysis Study. *Eurasian Journal of Educational Research*, 18, pp.1-22.

Paas, F.G., 1992. Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. *Journal of educational psychology*, *84*(4), p.429.

Paas, F., Renkl, A. and Sweller, J., 2003. Cognitive load theory and instructional design: Recent developments. *Educational psychologist*, 38(1), pp.1-4.

Paas, F., Van Gog, T. and Sweller, J., 2010. Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational psychology review*, 22(2), pp.115-121.

Paas, F. and van Merriënboer, J.J., 2020. Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science*, 29(4), pp.394-398.

Palincsar, A., 1998. Social Constructivist Perspectives on Teaching and Learning. *Annual Review of Psychology*, 49(1), pp.345-375.

Palmarini, R., Erkoyuncu, J., Roy, R. and Torabmostaedi, H., 2018. A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing*, 49, pp.215-228.

Parong, J. and Mayer, R.E., 2018. Learning science in immersive virtual reality. *Journal of Educational Psychology*, *110*(6), p.785.

Pedaste, M., Mäeots, M., Siiman, L., de Jong, T., van Riesen, S., Kamp, E., Manoli, C., Zacharia, Z. and Tsourlidaki, E., 2015. Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, pp.47-61.

Peleg, R., Yayon, M., Katchevich, D., Moria-Shipony, M. and Blonder, R., 2019. A lab-based chemical escape room: educational, mobile, and fun!. *Journal of chemical education*, *96*(5), pp.955-960.

Peña-Rios, A., Callaghan, V., Gardner, M. and Alhaddad, M.J., 2013, July. Developing xReality objects for mixed-reality environments. *In Intelligent Environments* (Workshops) (Vol. 17, pp. 190-200).

Pereshivkina, P., Karandasheva, N., Mikhaylenko, M. and Kurushkin, M., 2021. Immersive Molecular Dynamics in Virtual Reality: Increasing Efficiency of Educational Process with Companion Converter for NarupaXR. *Journal of Imaging*, *7*(6), p.97.

Perez, K.E., Strauss, E.A., Downey, N., Galbraith, A., Jeanne, R. and Cooper, S., 2010. Does displaying the class results affect student discussion during peer instruction?. *CBE—Life Sciences Education*, *9*(2), pp.133-140.

Perret, J. and Vander Poorten, E., 2018. Touching Virtual Reality: A Review of Haptic Gloves. *Conference: ACTUATOR 18At: Bremen, Germany*.

Peters, D., Calvo, R.A. and Ryan, R.M., 2018. Designing for motivation, engagement and wellbeing in digital experience. *Frontiers in psychology*, 9, p.797.

Pfennig, B.W. and Frock, R.L., 1999. The use of molecular modeling and VSEPR theory in the undergraduate curriculum to predict the three-dimensional structure of molecules. *Journal of chemical education*, 76(7), p.1018.

Pfundt, H., 1982. Pre-Instructional Conceptions about Transformations of Substances.

Piaget, J., 1970. Piaget's theory. In P. H. Mussen (Ed.), Carmichael's handbook of child psychology, pp.703-732. New York, NY: Wiley

Piaget, J., 1973. The affective unconscious and the cognitive unconscious. *Journal of the American Psychoanalytic Association*, *21*(2), pp.249-261.

Pienta, N.J., 2017. Studying student behavior and chemistry skill using browser-based tools and eye-tracking hardware. *Química Nova*, 40, pp.469-475.

Pietsch, J., Walker, R. and Chapman, E., 2003. The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school. *Journal of educational psychology*, 95(3), p.589.

Piñero Charlo, J.C., Ortega García, P. and Román García, S., 2021. Formative potential of the development and assessment of an educational escape room designed to integrate music-mathematical knowledge. *Education Sciences*, 11(3), p.131.

Plass, J.L., Homer, B.D. and Kinzer, C.K., 2015. Foundations of game-based learning. *Educational psychologist*, 50(4), pp.258-283.

Plunkett, K.N., 2019. A simple and practical method for incorporating augmented reality into the classroom and laboratory.

Pollock, E., Chandler, P. and Sweller, J., 2002. Assimilating complex information. *Learning and instruction*, *12*(1), pp.61-86.

Polvi, J., Taketomi, T., Noteki, A., Yoshitake, T., Fukuoka, T., Yamamoto, G., Sandor, C. and Kato, H., 2018. Handheld Guides in Inspection Tasks: Augmented Reality versus Picture. *IEEE Transactions on Visualization and Computer Graphics*, 24(7).

Porter, L., Bailey Lee, C., Simon, B. and Zingaro, D., 2011, August. Peer instruction: Do students really learn from peer discussion in computing?. In

*Proceedings of the seventh international workshop on Computing education research* (pp. 45-52).

Preacher, K.J., 2015. Extreme groups designs. *The encyclopedia of clinical psychology*, *2*, pp.1189-1192.

Prensky, M., 2003. Digital game-based learning. *Computers in Entertainment* (CIE), 1(1), pp.21-21.

Pribeanu, C., Balog, A. and Iordache, D.D., 2017. Measuring the perceived quality of an AR-based learning application: a multidimensional model. *Interactive Learning Environments*, *25*(4), pp.482-495.

Proulx, J.N., Romero, M. and Arnab, S., 2017. Learning mechanics and game mechanics under the perspective of self-determination theory to foster motivation in digital game based learning. *Simulation & Gaming*, 48(1), pp.81-97.

Pu, M. and Zhong, Z., 2018, May. Development of a situational interaction game for improving preschool children's performance in English-vocabulary learning. In Proceedings of the 2018 international conference on distance education and learning (pp. 88-92).

Qiao, X., Ren, P., Dustdar, S., Liu, L., Ma, H. and Chen, J., 2019. Web AR: A promising future for mobile augmented reality—State of the art, challenges, and insights. *Proceedings of the IEEE*, 107(4), pp.651-666.

Radu, I., 2012. Why should my students use AR? A comparative review of the educational impacts of augmented-reality. *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*.

Radu, I., 2014. Augmented reality in education: a meta-review and cross-media analysis. *Personal and ubiquitous computing*, 18(6), pp.1533-1543.

Ragan, E.D., Sowndararajan, A., Kopper, R. and Bowman, D.A., 2010. The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments. *Presence: Teleoperators and Virtual Environments*, *19*(6), pp.527-543.

Rahmawati, Y., Dianhar, H. and Arifin, F., 2021. Analysing students' spatial abilities in chemistry learning using 3D virtual representation. *Education Sciences*, *11*(4), p.185.

Raju, N.S., 1988. The area between two item characteristic curves. *Psychometrika*, *53*(4), pp.495-502.

Ramsden, P., 1997. The context of learning in academic departments. *The experience of learning*, 2, pp.198-216.

Rao, S.P. and DiCarlo, S.E., 2000. Peer instruction improves performance on quizzes. *Advances in physiology education*, 24(1), pp.51-55.

Rasimah, C.M.Y., Ahmad, A. and Zaman, H.B., 2011. Evaluation of user acceptance of mixed reality technology. *Australasian Journal of Educational Technology*, 27(8).

Ravna, O.V., Garcia, J., Themeli, C. and Prasolova-Førland, E., 2022. Supporting Peer-Learning with Augmented Reality in Neuroscience and Medical Education. In *KES International Conference on Smart Education and E-Learning* (pp. 299-308). Springer, Singapore.

Read, J.C. and Bohr, I., 2014. User experience while viewing stereoscopic 3D television. *Ergonomics*, 57(8), pp.1140-1153.

Reeve, J., 2012. A self-determination theory perspective on student engagement. In *Handbook of research on student engagement* (pp. 149-172). Springer, Boston, MA.

Reeve, J., Deci, E. L., and Ryan, R. M., 2004. Self-Determination Theory: A Dialectical Framework for Understanding Socio-Cultural Influences on Student Motivation. In D. M. Mclnerney, & S. Van Etten (Eds.), *Big Theories Revisited*, pp. 31-60. Greenwich, CT: Information Age Press.

Reisslein, J., Atkinson, R.K., Seeling, P. and Reisslein, M., 2006. Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and instruction*, 16(2), pp.92-103.

Renkl, A., Atkinson, R.K. and Große, C.S., 2004. How fading worked solution steps works–a cognitive load perspective. *Instructional science*, 32(1), pp.59-82.

Renninger, K. and Hidi, S., 2011. Revisiting the Conceptualization, Measurement, and Generation of Interest. *Educational Psychologist*, 46(3), pp.168-184.

Retnowati, E., 2018, March. Designing worked examples for learning tangent lines to circles. In *Journal of Physics: Conference Series* (Vol. 983, No. 1, p. 012124). IOP Publishing.

Rheinberg, F., Vollmeyer, R. and Burns, B.D., 2001. FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern-und Leistungssituationen (Langversion, 2001). *Diagnostica*, 2, pp.57-66.

Richard, E., Tijou, A., Richard, P. and Ferrier, J., 2006. Multi-modal virtual environments for education with haptic and olfactory feedback. *Virtual Reality*, 10(3-4), pp.207-225.

Richardson, M. and Abraham, C., 2009. Conscientiousness and achievement motivation predict performance. *European Journal of Personality*, 23(7), pp.589-605.

Richey, J.E. and Nokes-Malach, T.J., 2013. How much is too much? Learning and motivation effects of adding instructional explanations to worked examples. *Learning and Instruction*, 25, pp.104-124.

Roach, V.A., Mi, M., Mussell, J., Van Nuland, S.E., Lufler, R.S., DeVeau, K.M., Dunham, S.M., Husmann, P., Herriott, H.L., Edwards, D.N. and Doubleday, A.F., 2021. Correlating spatial ability with anatomy assessment performance: A meta-analysis. *Anatomical sciences education*, *14*(3), pp.317-329.

Rolnick, A. and Lubow, R.E., 1991. Why is the driver rarely motion sick? The role of controllability in motion sickness. *Ergonomics*, 34(7), pp.867-879.

Romli, R., Razak Yaakob, A., Abu Bakar, J., Syafiq Bahrin, A. and Nabila Zamalik, S., 2015. Design Features and Success Factors for Desktop VR in Learning Environment. *Journal of Computer Science & Computational Mathematics*, pp.49-53.

Rosenburg, L.B., 1992. The use of virtual fixtures as perceptual overlays to enhance operator performance in remote environments. In *Interim Report AL/CF-TR-1994-0089*. Armstrong Laboratory.

Rosenberg, M.J., Hovland, C.I., McGuire, W.J., Abelson, R.P. and Brehm, J.W., 1960. Attitude organization and change: An analysis of consistency among attitude components.(Yales studies in attitude and communication.), Vol. III.

Ross, R. and de Souza-Daw, A., 2021, April. Educational Escape Rooms as an Active Learning Tool for Teaching Telecommunications Engineering. In *Telecom* (Vol. 2, No. 2, pp. 155-166). MDPI.

Rourke, A. and Sweller, J., 2009. The worked-example effect using ill-defined problems: Learning to recognise designers' styles. *Learning and Instruction*, 19(2), pp.185-199.

Roy, R.V. and Zaman, B., 2017. Why gamification fails in education and how to make it successful: Introducing nine gamification heuristics based on self-determination theory. In *Serious Games and edutainment applications* (pp. 485-509). Springer, Cham.

Rubilar, M., Chova, L., Martínez, A. and Torres, I., 2019. Designing of augmented reality teaching-learning sequences to promote the accessibility and vizualization of complex contents in Chemistry. In *11th International Conference on Education and New Learning Technologies; IATED* (pp. 8299-8303).

Rutherford, F. and Ahlgren, A., 1990. *Science for all Americans*. New York: Oxford University Press.

Ryan, R.M. and Deci, E.L., 2020. Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary educational psychology*, 61, p.101860.

Ryan, R.M. and Deci, E.L., 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1), p.68.

Ryan, R.M. and Rigby, C.S., 2019. Motivational foundations of game-based learning. *Handbook of game-based learning*, pp.153-176.

Ryan, R.M., Ryan, W.S., Di Domenico, S.I. and Deci, E.L., 2019. The nature and the conditions of human autonomy and flourishing: Self-determination theory and basic psychological needs.

Safi, M., Chung, J. and Pradhan, P., 2019. Review of augmented reality in aerospace industry. *Aircraft Engineering and Aerospace Technology*.

Saidin, N.F., Halim, N.D.A. and Yahaya, N., 2019. Framework for Developing a Mobile Augmented Reality for Learning Chemical Bonds. *International Journal of Interactive Mobile Technologies*, 13(7).

Salomon, G., 1984. Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, 76(4), pp.647-658.

Saltan, F. and Arslan, Ö., 2016. The Use of Augmented Reality in Formal Education: A Scoping Review. *EURASIA Journal of Mathematics, Science and Technology Education*, 13(2).

Salzman, M.C., Dede, C., Loftin, R.B. and Chen, J., 1999. A model for understanding how virtual reality aids complex conceptual learning. *Presence: Teleoperators & Virtual Environments*, *8*(3), pp.293-316.

Sanii, B., 2019. Creating augmented reality USDZ files to visualize 3D objects on student phones in the classroom.

Santos, M.E.C., Taketomi, T., Yamamoto, G., Rodrigo, M., Mercedes, T., Sandor, C. and Kato, H., 2016. Augmented reality as multimedia: the case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning*, *11*(1), pp.1-23.

Sato, M., Liu, X., Murayama, J., Akahane, K. and Isshiki, M., 2008. A haptic virtual environment for molecular chemistry education. In *Transactions on edutainment I* (pp. 28-39). Springer, Berlin, Heidelberg.

Schechter, S., 2022. *What is markerless Augmented Reality?*. [online] 3D Cloud by Marxent. Available at: <https://www.marxentlabs.com/what-is-markerless-augmented-reality-dead-reckoning/> [Accessed 30 August 2022].

Schell, J. and Mazur, E., 2015. Flipping the chemistry classroom with peer instruction. *Chemistry Education: Best Practices, Opportunities and Trends*. Willey Online Library.

Schildkamp, K., van der Kleij, F.M., Heitink, M.C., Kippers, W.B. and Veldkamp, B.P., 2020. Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 103, p.101602.

Schnotz, W., 2010. Reanalyzing the expertise reversal effect. Instructional science, 38(3), pp.315-323.

Schroeder, N.L. and Cenkci, A.T., 2018. Spatial contiguity and spatial split-attention effects in multimedia learning environments: A meta-analysis. *Educational Psychology Review*, 30(3), pp.679-701.

Schunk, D., 2020. *Learning Theories: An Educational Perspective*. 8th ed. Pearson Education.

Schunk, D.H. and DiBenedetto, M.K., 2016. Self-efficacy theory in education. In Handbook of motivation at school (pp. 34-54). Routledge.

Schunk, D.H. and Pajares, F., 2005. Competence perceptions and academic functioning. Handbook of competence and motivation, 85, p.104.

Schunk, D.H. and Usher, E.L., 2012. Social cognitive theory and motivation. *The Oxford handbook of human motivation*, *2*, pp.11-26.

Schwab, J., 1962. The teaching of science as enquiry. In J. Schwab, & P. Brandwein (Eds.). The Teaching of Science, Cambridge, MA: Harvard University Press.

Scott, P., Mortimer, E. and Ametller, J., 2011. Pedagogical link-making: a fundamental aspect of teaching and learning scientific conceptual knowledge. *Studies in Science Education*, 47(1), pp.3-36.

SCRAP. A Party of Riddles, Puzzles, Curry and Beer (Translation), 2007. Available online at https://realdgame.jp/event/nazotokinoutage.html

Selfdeterminationtheory.org. 2022. *selfdeterminationtheory.org – Page Array – An approach to human motivation & personality*. [online] Available at: <https://selfdeterminationtheory.org/> [Accessed 14 September 2022].

Selzer, M., Gazcon, N. and Larrea, M., 2019. Effects of virtual presence and learning outcome using low-end virtual reality systems. *Displays*, 59, pp.9-15.

Sen, S., Yilmaz, A. and Temel, S., 2016. Adaptation of the Attitude toward the Subject of Chemistry Inventory (ASCI) into Turkish. *Journal of Education and Training Studies*, 4(8), pp.27-33.

Sepp, S., Howard, S.J., Tindall-Ford, S., Agostinho, S. and Paas, F., 2019. Cognitive load theory and human movement: Towards an integrated model of working memory. *Educational Psychology Review*, 31(2), pp.293-317.

Servotte, J., Goosse, M., Campbell, S., Dardenne, N., Pilote, B., Simoneau, I., Guillaume, M., Bragard, I. and Ghuysen, A., 2020. Virtual Reality Experience: Immersion, Sense of Presence, and Cybersickness. *Clinical Simulation in Nursing*, 38, pp.35-43.

Shaffer, D., Squire, K., Halverson, R. and Gee, J., 2005. Video Games and the Future of Learning. *Phi Delta Kappan*, 87(2), pp.105-111.

Shelton, B.E. and Hedley, N.R., 2004. Exploring a cognitive basis for learning spatial relationships with augmented reality. *Technology, Instruction, Cognition and Learning*, 1(4), p.323.

Shepard, R.N. and Cooper, L.A., 1986. *Mental images and their transformations*. The MIT Press.

Shin, D. and Biocca, F., 2018. Exploring immersive experience in journalism. *New media & society*, *20*(8), pp.2800-2823.

Sickel, J., 2019. The Great Media Debate and TPACK: A Multidisciplinary Examination of the Role of Technology in Teaching and Learning. *Journal of Research on Technology in Education*, 51(2), pp.152-165.

Silva, M., Roberto, R. and Teichrieb, V., 2015. Evaluation of Augmented Reality Technology in the English Language Field. *Anais do XXVI Simpósio Brasileiro de Informática na Educação* (SBIE 2015).

Simon, B., Kohanfars, M., Lee, J., Tamayo, K. and Cutts, Q., 2010, March. Experience report: peer instruction in introductory computing. In *Proceedings of the 41st ACM technical symposium on Computer science education* (pp. 341-345).

Sin, A.K. and Zaman, H.B., 2010, June. Live Solar System (LSS): Evaluation of an Augmented Reality book-based educational tool. In *2010 International Symposium on Information Technology* (Vol. 1, pp. 1-6). IEEE.

Singh, K., 2011. Study of achievement motivation in relation to academic achievement of students. *International Journal of Educational Planning & Administration*, 1(2), pp.161-171.

Sırakaya, M. and Alsancak Sırakaya, D., 2020. Augmented reality in STEM education: A systematic review. *Interactive Learning Environments*, pp.1-14.

Skarbez, R., Smith, M. and Whitton, M., 2021. Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality*, 2.

Smith, M.K., Wood, W.B., Adams, W.K., Wieman, C., Knight, J.K., Guild, N. and Su, T.T., 2009. Why peer discussion improves student performance on in-class concept questions. *Science*, *323*(5910), pp.122-124.

Smith, R. and Karaman, M.A., 2019. Development and validation of the contextual achievement motivation measure. *International Journal of Psychology and Educational Studies*, 6(3), pp.16-26.

Smith, R.L. and Troth, W.A., 1975. Achievement motivation: A rational approach to psychological education. *Journal of Counseling Psychology*, 22(6), p.500.

Snow, R.E., 1989. Aptitude-treatment interaction as a framework for research on individual differences in learning.

Sommerauer, P. and Müller, O., 2014. Augmented reality in informal learning environments: A field experiment in a mathematics exhibition. *Computers & education*, 79, pp.59-68.

Sommerauer, P. and Müller, O., 2018, June. Augmented Reality for Teaching and Learning-a literature Review on Theoretical and Empirical Foundations. In *ECIS* (p. 31).

Sorby, S., Drummer, T., Hungwe, K. and Charlesworth, P., 2005, June. Developing 3 D spatial visualization skills for non engineering students. In *2005 Annual Conference* (pp. 10-428).

Sørebø, Ø., Halvari, H., Gulli, V.F. and Kristiansen, R., 2009. The role of self-determination theory in explaining teachers' motivation to continue to use e-learning technology. *Computers & Education*, 53(4), pp.1177-1187.

Southgate, E. and Smith, S., 2016. Pedagogical theory and embodiment: Some provocations for virtual and augmented reality education. In *IEEE VR Fifth*

*Workshop on K-12+ Embodied Learning through Virtual & Augmented Reality (KELVAR) IEEVR*.

Stajkovic, A.D., Lee, D. and Nyberg, A.J., 2009. Collective efficacy, group potency, and group performance: meta-analyses of their relationships, and test of a mediation model. *Journal of applied psychology*, *94*(3), p.814.

Stafford, R.E., 1961. Sex differences in spatial visualization as evidence of sex-linked inheritance. *Perceptual and motor skills*, *13*(3), pp.428-428.

Steiner, R. and Sullivan, J., 1984. Variables correlating with student success in organic chemistry. *Journal of Chemical education*, 61(12), p.1072.

Stieff, M., 2007. Mental rotation and diagrammatic reasoning in science. *Learning and instruction*, 17(2), pp.219-234.

Stoffregen, T.A., Chen, Y.C. and Koslucher, F.C., 2014. Motion control, motion sickness, and the postural dynamics of mobile devices. *Experimental brain research*, *232*(4), pp.1389-1397.

Strack, J., Lopes, P., Esteves, F. and Fernandez-Berrocal, P., 2017. Must we suffer to succeed? When anxiety boosts motivation and performance. *Journal of Individual Differences*, 38(2), p.113.

Suits, J.P. and Sanger, M.J., 2013. Dynamic visualizations in chemistry courses. In *Pedagogic roles of animations and simulations in chemistry courses* (pp. 1-13). American Chemical Society.

Sutherland, I.E., 1968, December. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I* (pp. 757-764).

Sutherland, L.M., Middleton, P.F., Anthony, A., Hamdorf, J., Cregan, P., Scott, D. and Maddern, G.J., 2006. Surgical simulation: a systematic review. *Annals of surgery*, *243*(3), p.291.

Swarat, S., Ortony, A. and Revelle, W., 2012. Activity matters: Understanding student interest in school science. *Journal of Research in Science Teaching*, 49(4), pp.515-537.

Sweeder, R.D., Herrington, D.G. and VandenPlas, J.R., 2019. Supporting students' conceptual understanding of kinetics using screencasts and simulations outside of the classroom. *Chemistry Education Research and Practice*, 20(4), pp.685-698.

Sweller, J., 1988. Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2), pp.257-285.

Sweller, J., 2006. The worked example effect and human cognition. *Learning and instruction*.

Sweller, J. and Chandler, P., 1994. Why some material is difficult to learn. *Cognition and instruction*, 12(3), pp.185-233.

Sweller, J. and Cooper, G.A., 1985. The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and instruction*, 2(1), pp.59-89.

Sweller, J., Ayres, P. and Kalyuga, S., 2011. Amassing information: The information store principle. In *Cognitive load theory* (pp. 17-25). Springer, New York, NY.

Sweller, J., Ayres, P.L., Kalyuga, S. and Chandler, P., 2003. The expertise reversal effect.

Sweller, J., 2011. Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press.

Sweller, J., Van Merriënboer, J.J. and Paas, F.G., 1998. Cognitive architecture and instructional design. *Educational psychology review*, 10(3), pp.251-296.

Sweller, J., van Merriënboer, J.J. and Paas, F., 2019. Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), pp.261-292.

Taajamaa, V., Jarvi, A., Laato, S. and Holvitie, J., 2018. Co-creative engineering curriculum design — Case East Africa. *2018 IEEE Frontiers in Education Conference (FIE)*.

Taber, K., 2002. *Chemical misconceptions: prevention, diagnosis and cure* (Vol. 1). Royal Society of Chemistry.

Tanaka, A. and Yamauchi, H., 2001. A model for achievement motives, goal orientations, intrinsic interest, and academic achievement. *Psychological Reports*, 88(1), pp.123-135.

Tang, H. and Pienta, N., 2012. Eye-tracking study of complexity in gas law problems. *Journal of Chemical Education*, 89(8), pp.988-994.

Taylor, G., Jungert, T., Mageau, G.A., Schattke, K., Dedic, H., Rosenfield, S. and Koestner, R., 2014. A self-determination theory approach to predicting school achievement over time: The unique role of intrinsic motivation. *Contemporary educational psychology*, 39(4), pp.342-358.

Tee, N., Gan, H., Li, J., Cheong, B., Tan, H., Liew, O. and Ng, T., 2018. Developing and Demonstrating an Augmented Reality Colorimetric Titration Tool. *Journal of Chemical Education*, 95(3), pp.393-399.

ter Vrugte, J., de Jong, T., Vandercruysse, S., Wouters, P., van Oostendorp, H. and Elen, J., 2017. Computer game-based mathematics education: Embedded faded worked examples facilitate knowledge acquisition. *Learning and instruction*, 50, pp.44-53.

Tercanli, H., Martina, R., Ferreira Dias, M., Wakkee, I., Reuter, J., Amorim, M., Madaleno, M., Magueta, D., Vieira, E., Veloso, C. and Figueiredo, C., 2021. Educational escape rooms in practice: research, experiences, and recommendations.

Theobald, R. and Freeman, S., 2014. Is it the intervention or the students? Using linear regression to control for student characteristics in undergraduate STEM education research. *CBE—Life Sciences Education*, 13(1), pp.41-48.

Theodoropoulos, A. and Lepouras, G., 2021. Augmented Reality and programming education: A systematic review. *International Journal of Child-Computer Interaction*, *30*, p.100335.

Tindall-Ford, S., Chandler, P. and Sweller, J., 1997. When two sensory modes are better than one. *Journal of experimental psychology*: Applied, 3(4), p.257.

Trujillo, G. and Tanner, K.D., 2014. Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity. *CBE—Life Sciences Education*, 13(1), pp.6-15.

Tsigilis, N. and Theodosiou, A., 2003. Temporal stability of the intrinsic motivation inventory. *Perceptual and motor skills*, 97(1), pp.271-280.

Tuckey, H., Selvaratnam, M. and Bradley, J., 1991. Identification and rectification of student difficulties concerning three-dimensional structures, rotation, and reflection. *Journal of Chemical Education*, 68(6), p.460.

Tuminaro, J. and Redish, E.F., 2007. Elements of a cognitive model of physics problem solving: Epistemic games. *Physical Review Special Topics-Physics Education Research*, 3(2), p.020101.

Turan, Z., Avinc, Z., Kara, K. and Goktas, Y., 2016. Gamification and education: Achievements, cognitive loads, and views of students. *International journal of emerging technologies in learning*, *11*(7).

Turan, Z., Meral, E. and Sahin, I., 2018. The impact of mobile augmented reality in geography education: achievements, cognitive loads and views of university students. *Journal of Geography in Higher Education*, 42(3), pp.427-441.

Tüysüz, C., 2010. The Effect of the Virtual Laboratory on Students' Achievement and Attitude in Chemistry. *International Online Journal of Educational Sciences*, *2*(1).

Ugi, I., Marquarding, D., Klusacek, H., Gillespie, P. and Ramirez, F., 1971. Berry pseudorotation and turnstile rotation. *Accounts of Chemical Research*, 4(8), pp.288-296.

Um, E., Plass, J.L., Hayward, E.O. and Homer, B.D., 2012. Emotional design in multimedia learning. *Journal of educational psychology*, 104(2), p.485.

Unity., 2022. Unity. [online] Unity. Available at: <https://unity.com/> [Accessed 11 September 2022].

Unity3d., 2022. Unity - Scripting API: Collider. [online] Docs.unity3d.com. Available at: <https://docs.unity3d.com/ScriptReference/Collider.html> [Accessed 11 September 2022].

Urhahne, D., Nick, S. and Schanze, S., 2009. The effect of three-dimensional simulations on the understanding of chemical structures and their properties. *Research in science education*, 39(4), pp.495-513.

Vallerand, R.J., Fortier, M.S. and Guay, F., 1997. Self-determination and persistence in a real-life setting: toward a motivational model of high school dropout. *Journal of Personality and Social psychology*, 72(5), p.1161.

Van den Akker, J., Gravemeijer, K. and McKenney, S., 2006. Introducing educational design research. In *Educational design research* (pp. 15-19). Routledge.

Van Gog, T., Kester, L. and Paas, F., 2011. Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, 36(3), pp.212-218.

Van Gerven, P.W., Paas, F.G., Van Merriënboer, J.J. and Schmidt, H.G., 2002. Cognitive load theory and aging: Effects of worked examples on training efficiency. *Learning and instruction*, 12(1), pp.87-105.

Van Haren, R., 2010. Engaging learner diversity through learning by design. *E-learning and Digital Media*, 7(3), pp.258-271.

Van Merriënboer, J.J., Kirschner, P.A. and Kester, L., 2003. Taking the load off a learner's mind: Instructional design for complex learning. *Educational psychologist*, 38(1), pp.5-13.

Van Merriënboer, J.J. and Sweller, J., 2005. Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review*, 17(2), pp.147-177.

VandenPlas, J.R., Herrington, D.G., Shrode, A.D. and Sweeder, R.D., 2021. Use of simulations and screencasts to increase student understanding of energy concepts in bonding. *Journal of Chemical Education*, 98(3), pp.730-744.

Vandenberg, S.G., Kuse, A.R. and Vogler, G.P., 1985. Searching for correlates of spatial ability. *Perceptual and motor skills*, *60*(2), pp.343-350.

Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K. and Lens, W., 2009. Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of educational psychology*, 101(3), p.671.

Veidis, M.V., Schreiber, G.H., Gough, T.E. and Palenik, G.J., 1969. Jahn-Teller distortions in octahedral copper (II) complexes. *Journal of the American Chemical Society*, 91(7), pp.1859-1860.

Vergauwe, E., Camos, V. and Barrouillet, P., 2014. The impact of storage on processing: How is information maintained in working memory?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(4), p.1072.

Vergne, M.J., Simmons, J.D. and Bowen, R.S., 2019. Escape the lab: An interactive escape-room game as a laboratory experiment. *Journal of Chemical Education*, 96(5), pp.985-991.

Vergne, M.J., Smith, J.D. and Bowen, R.S., 2020. Escape the (remote) classroom: An online escape room for remote learning. *Journal of chemical education*, 97(9), pp.2845-2848.

Vicari, C., 2020. Escape the Planet: Empowering Student Designers to Create a Science-Based Escape Room with Augmented Reality. *International Journal of Designs for Learning*, 11(2), pp.80-95.

Vickrey, T., Rosploch, K., Rahmanian, R., Pilarz, M. and Stains, M., 2015. Research-based implementation of peer instruction: A literature review. *CBE—Life Sciences Education*, 14(1), p.es3.

Vidergor, H.E., 2021. Effects of digital escape room on gameful experience, collaboration, and motivation of elementary school students. *Computers & Education*, 166, p.104156.

Vienne, C., Masfrand, S., Bourdin, C. and Vercher, J.L., 2020. Depth perception in virtual reality systems: Effect of screen distance, environment richness and display factors. *IEEE Access*, *8*, pp.29099-29110.

Virata, R.O. and Castro, J.D.L., 2019, January. Augmented reality in science classroom: Perceived effects in education, visualization and information processing. In *Proceedings of the 10th International Conference on E-Education, E-Business, E-Management and E-Learning* (pp. 85-92).

Visual Studio. 2022. Visual Studio: IDE and Code Editor for Software Developers and Teams. [online] Available at: <https://visualstudio.microsoft.com/> [Accessed 11 September 2022].

Vollmeyer, R. and Rheinberg, F., 2006. Motivational effects on self-regulated learning with different tasks. *Educational Psychology Review*, 18(3), pp.239-253.

Vörös, A.I.V. and Sárközi, Z., 2017, December. Physics escape room as an educational tool. In *AIP Conference Proceedings* (Vol. 1916, No. 1, p. 050002). AIP Publishing LLC.

Vrabec, M. and Prokša, M., 2016. Identifying Misconceptions Related to Chemical Bonding Concepts in the Slovak School System Using the Bonding Representations Inventory as a Diagnostic Tool. *Journal of Chemical Education*, 93(8), pp.1364-1370.

Vygotsky, L., 1978. Interaction between learning and development. *Readings on the development of children*, 23(3), pp.34-41.

Vygotsky, L., Cole, M., Stein, S. and Sekula, A., 1978. *Mind in society*. Cambridge, Mass.: Harvard University Press.

Wadt, W.R. and Hay, P.J., 1985. Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi. *The Journal of chemical physics*, 82(1), pp.284-298.

Wai, J., Lubinski, D. and Benbow, C.P., 2009. Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of educational Psychology*, 101(4), p.817.

Ware, C. and Rose, J., 1999. Rotating virtual objects with real handles. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 6(2), pp.162-180.

Weinbaum, S., 1935. *Pygmalion's Spectacles*. Independently Published.

Wen, Y. and Looi, C., 2019. Review of Augmented Reality in Education: Situated Learning with Digital and Non-digital Resources. *Smart Computing and Intelligence*, pp.179-193.

Wentzel, K.R., 2020. *Motivating students to learn*. Routledge.

Wieman, C., 2017. *Improving How Universities Teach Science*. Cambridge: Harvard Univ Press.

Williams, P., 2018, March. Using escape room-like puzzles to teach undergraduate students effective and efficient group process skills. In *2018 IEEE Integrated STEM Education Conference (ISEC)* (pp. 254-257). IEEE.

Williamson, V., Hegarty, M., Deslongchamps, G., Williamson, K. and Shultz, M., 2013. Identifying Student Use of Ball-and-Stick Images versus Electrostatic Potential Map Images via Eye Tracking. *Journal of Chemical Education*, 90(2), pp.159-164.

Winn, W. and Jackson, R., 1999. Fourteen propositions about educational uses of virtual reality. *Educational Technology*, 39(4), pp.5-14.

Wojciechowski, R. and Cellary, W., 2013. Evaluation of learners' attitude toward learning in ARIES augmented reality environments. *Computers & Education*, 68, pp.570-585.

Wong, C.H., Tsang, K.C. and Chiu, W.K., 2021. Using Augmented Reality as a Powerful and Innovative Technology to Increase Enthusiasm and Enhance Student Learning in Higher Education Chemistry Courses. *Journal of Chemical Education*, 98(11), pp.3476-3485.

Wood, J. and Donnelly-Hermosillo, D.F., 2019. Learning chemistry nomenclature: Comparing the use of an electronic game versus a study guide approach. *Computers & Education*, 141, p.103615.

Wu, B., Yu, X. and Gu, X., 2020. Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, *51*(6), pp.1991-2005.

Wu, W.H., Hsiao, H.C., Wu, P.L., Lin, C.H. and Huang, S.H., 2012. Investigating the learning-theory foundations of game-based learning: a meta-analysis. *Journal of Computer Assisted Learning*, 28(3), pp.265-279.

Wu, H.K. and Shah, P., 2004. Exploring visuospatial thinking in chemistry learning. *Science education*, 88(3), pp.465-492.

Wulandari, I., Irwansyah, F., Farida, I. and Ramdhani, M., 2019. Development of student's submicroscopic representation ability on molecular geometry material using Augmented Reality (AR) media. *Journal of Physics: Conference Series*, 1280(3), p.032016.

Xiao, J., Xu, Z., Yu, Y., Cai, S. and Hansen, P., 2016, April. The design of augmented reality-based learning system applied in U-learning environment. In *International conference on technologies for E-learning and digital entertainment* (pp. 27-36). Springer, Cham.

Xu, X. and Lewis, J.E., 2011. Refinement of a chemistry attitude measure for college students. *Journal of Chemical Education*, 88(5), pp.561-568.

Xu, X., Villafane, S.M. and Lewis, J.E., 2013. College students' attitudes toward chemistry, conceptual knowledge and achievement: structural equation model analysis. *Chemistry Education Research and Practice*, 14(2), pp.188-200.

Yang, E.M., Andre, T., Greenbowe, T.J. and Tibell, L., 2003. Spatial ability and the impact of visualization/animation on learning electrochemistry. *International Journal of Science Education*, 25(3), pp.329-349.

Yang, S., Mei, B. and Yue, X., 2018. Mobile augmented reality assisted chemical education: insights from elements 4D.

Yilmaz, K., 2008. Constructivism: Its Theoretical Underpinnings, Variations, and Implications for Classroom Instruction. *Educational Horizons*, *86*(3), pp.161–172.

Yu, D., Jin, J.S., Luo, S., Lai, W. and Huang, Q., 2009. A useful visualization technique: A literature review for augmented reality and its application, limitation & future direction. *Visual information communication*, pp.311-337.

Zajacova, A., Lynch, S.M. and Espenshade, T.J., 2005. Self-efficacy, stress, and academic success in college. *Research in higher education*, 46(6), pp.677-706.

Zaug, P., Gros, C.I., Wagner, D., Pilavyan, E., Meyer, F., Offner, D. and Strub, M., 2022. Development of an innovative educational escape game to promote teamwork in dentistry. *European Journal of Dental Education*, 26(1), pp.116-122.

Zeng, H., He, X. and Pan, H., 2021. Implementation of escape room system based on augmented reality involving deep convolutional neural network. *Virtual Reality*, 25(3), pp.585-596.

Zhang, J., Liu, T., Sung, Y. and Chang, K., 2015. Using Augmented Reality to Promote Homogeneity in Learning Achievement. *2015 IEEE International Symposium on Mixed and Augmented Reality - Media, Art, Social Science, Humanities and Design*.

Zhu, B., Feng, M., Lowe, H., Kesselman, J., Harrison, L. and Dempski, R.E., 2018. Increasing enthusiasm and enhancing learning for biochemistry-laboratory safety with an augmented-reality program. *Journal of Chemical Education*, 95(10), pp.1747-1754.

# Appendix A – Simulation Sickness Questionnaire

No______________ Date____________ Pre/Post_________

Instructions: Circle how much each symptom below is affecting you right now.

1. General discomfort

None Slight Moderate Severe

2. Fatigue

None Slight Moderate Severe

3. Headache

None Slight Moderate Severe

4. Eye strain

None Slight Moderate Severe

5. Difficulty focusing

None Slight Moderate Severe

6. Salivation increasing

None Slight Moderate Severe

7. Sweating

None Slight Moderate Severe

8. Nausea

None Slight Moderate Severe

9. Difficulty concentrating

None Slight Moderate Severe

10. « Fullness of the Head »

| | | | |
|---|---|---|---|
| None | Slight | Moderate | Severe |

11. Blurred vision

| | | | |
|---|---|---|---|
| None | Slight | Moderate | Severe |

12. Dizziness with eyes open

| | | | |
|---|---|---|---|
| None | Slight | Moderate | Severe |

13. Dizziness with eyes closed

| | | | |
|---|---|---|---|
| None | Slight | Moderate | Severe |

14. *Vertigo

| | | | |
|---|---|---|---|
| None | Slight | Moderate | Severe |

15. **Stomach awareness

| | | | |
|---|---|---|---|
| None | Slight | Moderate | Severe |

16. Burping

| | | | |
|---|---|---|---|
| None | Slight | Moderate | Severe |

* Vertigo is experienced as loss of orientation with respect to vertical upright.

** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea.

# Appendix B – Interview Schedule for EEA 2020

**Question 1:** How interested were you to learn the topics within the presented learning activity?

**Question 2:** How did this *Augmented Reality* technology help your understanding of *VSEPR*?

**Question 3:** Were you satisfied with your own performance when using the *Augmented Reality* technology?

**Question 4:** How did this *Augmented Reality* technology help you to analyse problems presented to you around the topic?

**Question 5:** What were the best / worst aspects of your experience using the *Augmented Reality* technology?

**Question 6:** What improvements would you make to the *Augmented Reality* tool?

**Question 7:** Would you recommend this *Augmented Reality* technology to other UEA students and staff members? Why?

# Appendix C – ITMC test Instrument

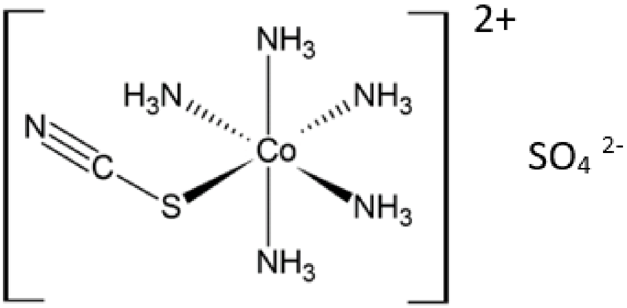
1. What is the systematic name for the complex anion $[NiCl_4]^{2-}$?

    a. Tetrachloridonickel(II)
    b. Tetrachloridonickelate(II)
    c. Tetrakischloridonickelate(II)
    d. Tetrakischloridenickel(II) ion

2. A coordination compound contains both fluorido ($F^-$) and ammine ($NH_3$) ligands. Which of the two ligands would be expected to appear first in the compound's name?

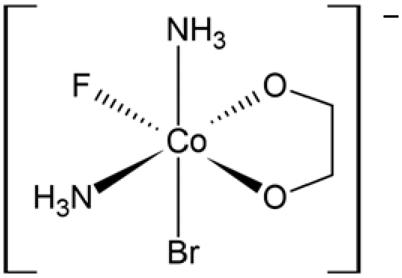
    a. Ammine, because it is a neutral ligand
    b. Ammine, because of alphabetical ordering
    c. Fluorido, because it is an anionic ligand
    d. Fluorido, because it has a higher atomic mass than N

3. What is the systematic name of the following linkage isomer?




    a. pentaamminethiocyanato-κ*S*-cobalt(III) sulfate
    b. pentaamminethiocyanato-κ*N*-cobalt(II) sulfate
    c. pentaamminethiocyanato-κ*S*-cobalt(II) sulfate
    d. pentaamminethiocyanato-κ*N*-cobalt(III) sulfate

4. For the following isomer, which correctly describes the relationship of the ligands?

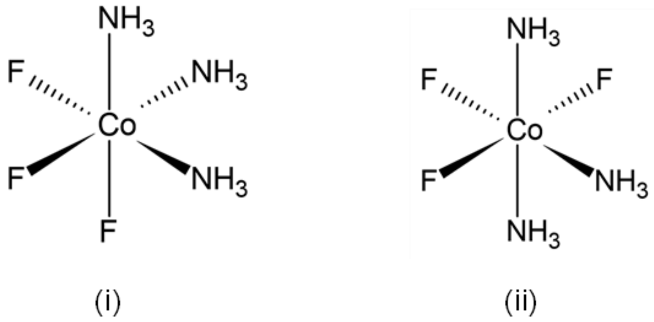



a. $NH_3$ is *trans* to $NH_3$, and Br is *cis* to F
b. $NH_3$ is *cis* to $NH_3$, and Br is *trans* to F
c. $NH_3$ is *trans* to $NH_3$, and Br is *trans* to F
d. $NH_3$ is *cis* to $NH_3$, and Br is *cis* to F

5. How many geometric isomers are possible for a square planar complex of general formula $[MA_2B_2]$?

a. 0
b. 1
c. 2
d. 3

6. Two isomers of coordination compound $[Co(NH_3)_3F_3]$ are shown below. The isomers can be classified as:




a. (i) fac-isomer (ii) mer-isomer
b. (i) optical isomer (ii) trans-isomer
c. (i) mer-isomer (ii) fac-isomer
d. (i) trans-isomer (ii) cis-isomer

7. What type of isomerism is displayed by complexes $[Co(NH_3)_6][Cr(CN)_6]$ and $[Co(CN)_6][Cr(NH_3)_6]$?

   a. Coordination isomerism
   b. Hydrate isomerism
   c. Ionisation isomerism
   d. Linkage isomerism

8. The coordination compounds with the composition $CrCl_3.6H_2O$ exists as different isomers displaying a range of colours including violet and green. What isomerism is this a result of?

   a. Ionisation isomerism
   b. Coordination isomerism
   c. Optical isomerism
   d. Hydrate isomerism

9. **Statement (i):** Linkage isomerism can occur in coordination compounds containing ambidentate ligands.
   **Statement (ii):** Ambidentate ligands have more than one potential donor atom

   a. **Both statements (i) and (ii) are correct, and (ii) explains (i).**
   b. Both statements (i) and (ii) are correct, but (ii) does not explain (i).
   c. Statement (i) is correct, but (ii) is incorrect.
   d. Both statements (i) and (ii) are incorrect.

10. Match each isomer pair on the left with the correct type of isomerism on the right.

| **Isomer Pair** | | **Isomerism** | |
|---|---|---|---|
| 1. | $[Co(NH_3)_5(NO_2)]Cl_2$ and $[Co(NH_3)_5(ONO)]Cl_2$ | A. | Ionisation |
| 2. | $[Cu(NH_3)_4][PtCl_4]$ and $[Pt(NH_3)_4][CuCl_4]$ | B. | Linkage |
| 3. | $[PtCl_2(NH_3)_4]Br_2$ and $[PtBr_2(NH_3)_4]Cl_2$ | C. | Coordination |

# Appendix D – Interview Schedule for EEA 2021/2022

**Perception/Satisfaction.**

**Question 1:** Could you tell me about your levels of satisfaction regarding this type of chemistry learning experience?

> **Extension**: Were you satisfied with your own performance?
>
> **Extension:** Did you find the experience fun/engaging?
>
> **Extension:** Would you recommend this method of learning to your peers?

**Question 2:** Did you identify any shortcomings or challenges when carrying out this learning experience. What can be improved?

**GBL paradigm.**

**Question 3:** What is your interest/experience in gaming? (Whether this is within/outside education)

> (*Plays computer games)*
>
> **Extension:** How many hours per week do you play games?
>
> **Extension**: What stimulates you to keep on playing a game when it gets progressively more difficult?
>
> **Extension:** Do you prefer to play games individually or with others? (Cooperatively or competitively.)
>
> *(Doesn't play games)*
>
> **Extension:** If you had an opportunity to use games for learning in your degree programme, is this a learning experience that you believe could be meaningful?
>
> > **Extension:** If not, why is this the case?

**Value/Usefulness.**

**Question 4:** How beneficial do you believe this activity was to you as a student? (Was it meaningful?)

**Question 5:** Do you believe this activity was useful for evaluating your understanding of stereochemistry concepts?

> **Extension**: Does this activity provide the opportunity for you to better apply what you had learned?

**Extension**: Were you able to reflect on your own understanding? (Will the game help you remember what was learned?)

**Question 6:** Would you be willing to carry out this style of activity again?

**Pressure/Tension and Effort/Importance.**

**Question 7:** How much effort did you put into this activity? Was it important for you to do well at this task?

**Extension:** Did this influence your learning satisfaction?

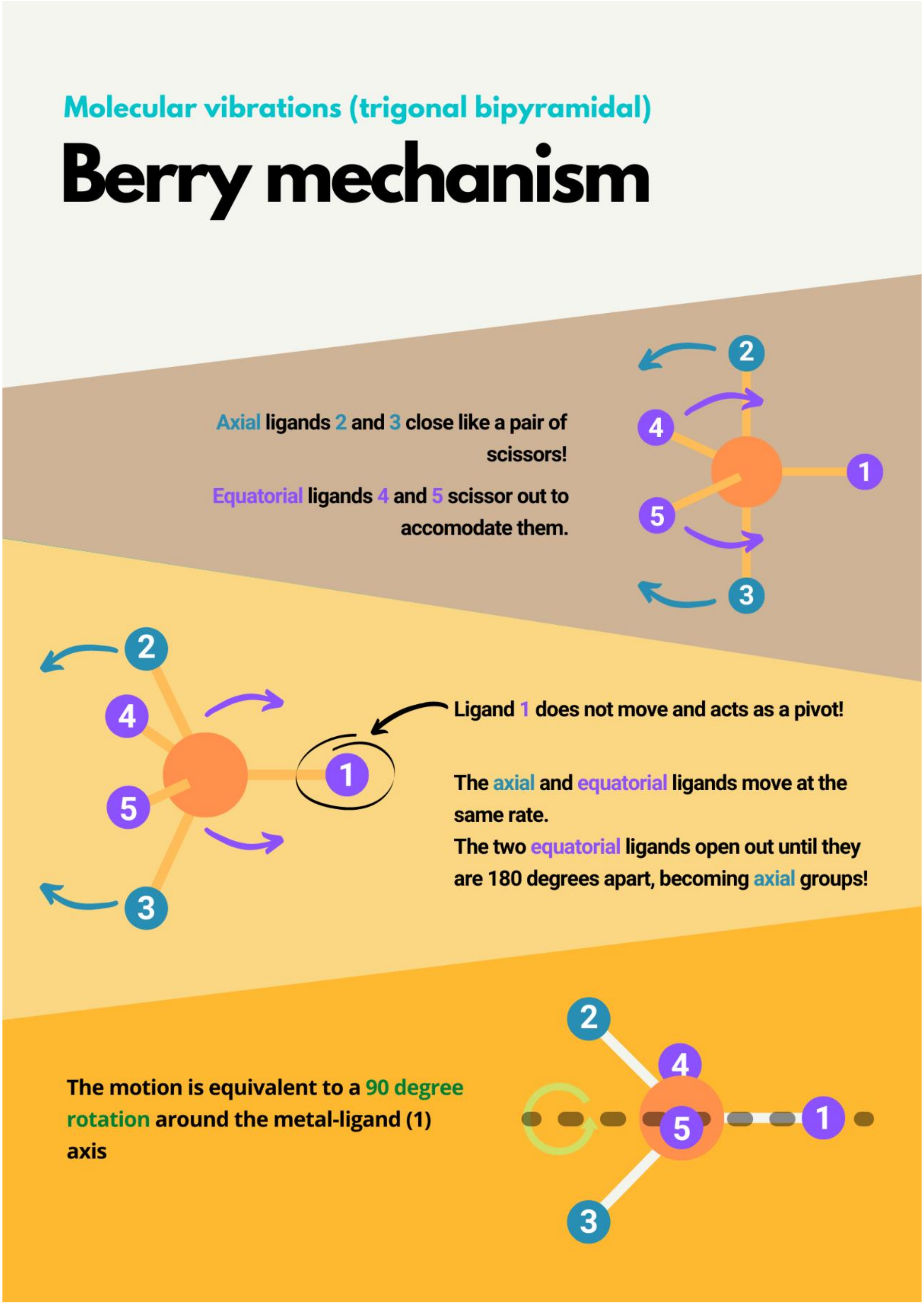
**Question 8:** Were you nervous/anxious whilst carrying out this activity?

**Extension:** Did you feel pressured whilst carrying out the activity?

# Appendix E – Negotiated codebook for EEA 2021/2022

| Code | Code Description |
|---|---|
| 1. Negative student feelings | Negative views from students (independent of the learning experience) |
| 1a. Anxiety | Feelings of anxiety expressed by students |
| 1b. Inactivity | "Dead zones" throughout the activity experience |
| 2. Learning | Comments on the learning experience |
| 2a. Application of the subject content | Application of the stereochemistry subject content within the activity |
| 2b. Discussion of chemistry | Discussion of chemistry (not exclusively stereochemistry) |
| 2c. Use of existing resources | Discussion of other resources utilized by students |
| 2d. Reflection on the learning process | To include references to revision |
| 2e. Activity motivation | What drives players to continue THIS activity when encountering a difficult problem? |
| 2f. Augmented reality (AR) | Comments on the AR technology |
| 3. Completion rate | Statements regarding activity completion rate |
| 4. Future version | Comments on general Education Escape Activity development (including AR integration) |
| 4a. Technical fixes | Any technical bugs/issues that require attention/fixing |
| 4b. Developing the activity | Suggestions on developing THIS activity |
| 5. Embedding Game-Based Learning (GBL) | Statements on embedding GBL elements into the teaching activity |
| 5a. Puzzle dynamic | Views on the puzzle mechanics within the activity (difficulty, construction, etc.) |
| 5b. Chemistry problems | Views on how the subject content has been integrated into the puzzles |
| 5c. Teamwork | To include collaboration in solving chemistry problems |
| 6. Prior gaming experience | Experience/interest in gaming outside of formal education |
| 6a. Cooperative/competitive | How do participants engage in gaming as an interest? |
| 6b. Gaming motivation | What drives players to continue playing when a game becomes challenging? |
| 7. Positive student feelings | Positive views from students (independent of the learning experience) |
| 8. Social interactivity | Students interacting throughout the activity |
| 8a. Communication | Student discussion within their team |
| 8b. Inclusion | Challenges that may hinder inclusion of students |

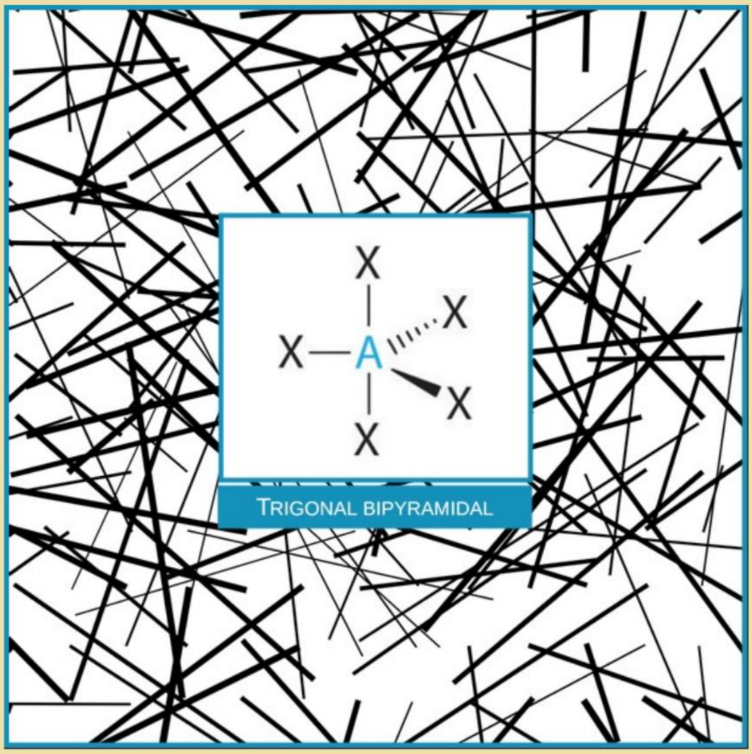
# Appendix F – Pretraining exercise



Molecular vibrations (trigonal bipyramidal)

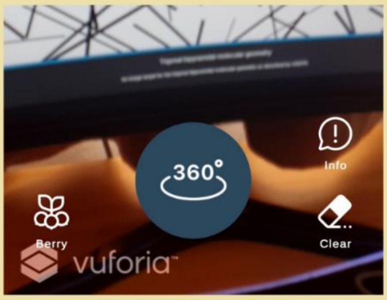
# Berry mechanism

Scan me using ChemFord!

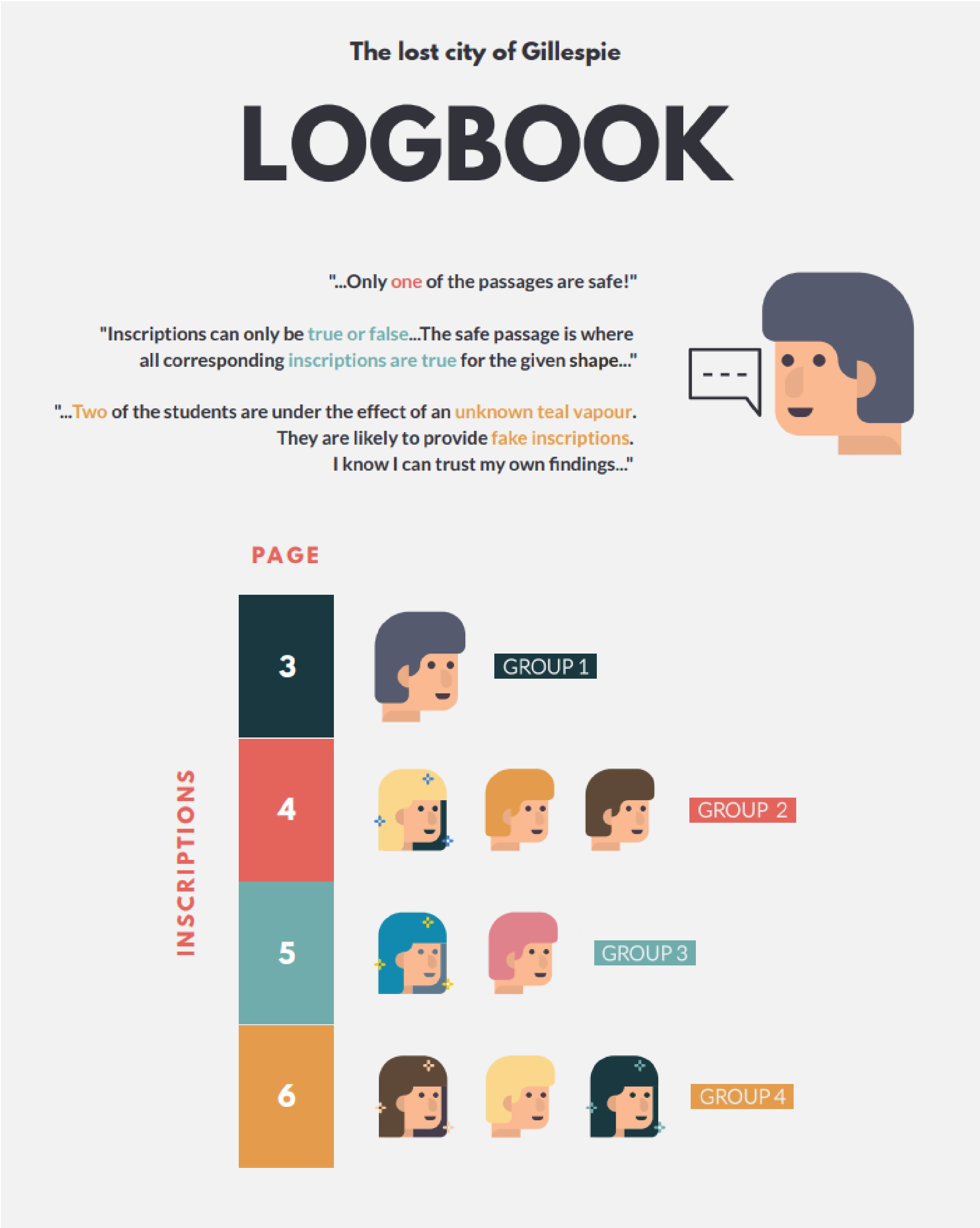


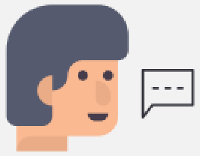

**STEP 1**
Scan the above image in ChemFord to generate the trigonal bipyramidal geometry (try moving it with your finger!)

**STEP 2**
Press the "Berry" button to see the mechanism in action!



# Appendix G – VSEPR Intervention

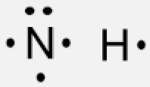




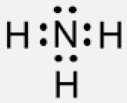
The analysis of the inscriptions is likely to involve the generation of dash-wedge structrures. One approach I could use is shown below!

$NH_3$

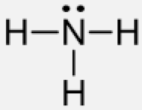
1. Write the chemical formula



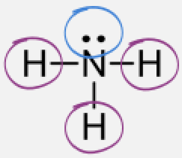
2. Draw the Lewis dot structure of each atom



3. Draw the Lewis structure based on the atoms drawn

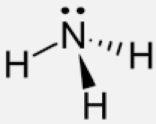


4. Redraw the structure using solid lines



Steric number = 3 + 1

5. Deduce the steric number (Bound atoms + lone pairs)



Trigonal pyramidal

$AX_3E$

6. Draw the correct geometry based on the steric number

# Passage 1 Inscriptions

GROUP 1

True (T) / False (F)

Scan me with ChemFord

1. $H_2O$ and $SCl_2$ both adopt this geometry and exhibit dipole moments... T / F

2. This geometry is known as bent or angular... T / F

3. Substituting S in $SCl_2$ with Be retains both the geometry and dipole moment... T / F

4. The lone pairs repel less strongly than the bonding groups in this geometry... T / F

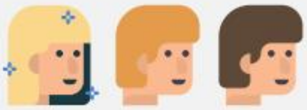
5. This geometry can have a steric number of three or four... T / F

6. Ozone and $SO_2$ do not adopt this molecular geometry as their double bonds contribute more than one bonding group... T / F

7. The bond angles in this geometry are less than those in a linear geometry... T / F

8. Addition of a bonding group would result in a trigonal pyramidal geometry if one lone pair was present... T / F

# Passage 2 Inscriptions



GROUP 2

True (T) / False (F)

Scan me with ChemFord

1. The equatorial bound groups are separated by a bond angle of 120º... T / F

2. The equatorial and axial bound groups are separated by a bond angle of 90º... T / F

3. The equatorial and axial groups are not equivalent... T / F

4. Within this geometry, there are five bonding electron regions and no lone pairs... T / F

5. This geometry exhibits Berry pseudorotation... T / F

6. Replacing an equatorial bound group with a lone pair in this geometry would result in the seesaw geometry... T / F

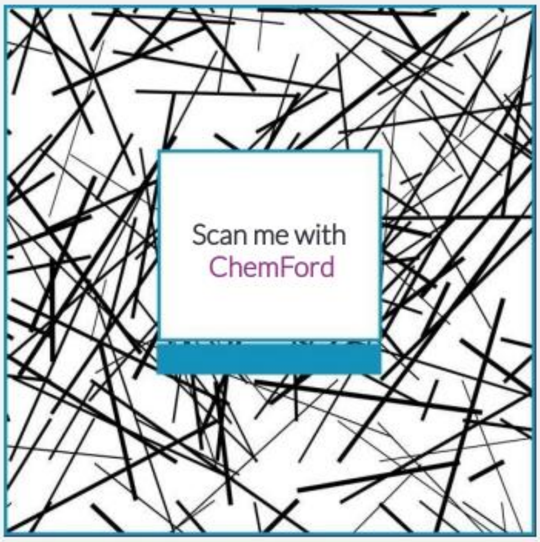
7. The two axial bound groups are separated by a bond angle of 180º... T / F

8. This geometry is called trigonal bipyramidal... T / F

9. $PF_5$ adopts this geometry, as does $Fe(CO)_5$... T / F

# Passage 3 Inscriptions

True (T) / False (F)




1. $CH_4$ and $CF_4$ both adopt this geometry and are perfectly symmetrical... T / F

2. The bond angle of this geometry is 109.5º... T / F

3. 100% symmetrical molecules of this geometry are always non-polar... T / F

4. Square planar and seesaw geometries have the same number of bonding groups as this geometry... T / F

5. This geometry has a steric number of four and no lone pairs... T / F

6. Berry pseudorotation is not observed in molecules adopting this geometry... T / F

7. This geometry is known as tetrahedral... T / F

8. $CH_3Cl$ adopts this geometry and displays a dipole moment... T / F

9. The phosphate ion adopts this geometry... T / F

10. The shape of this geometry is based on a tetrahedron... T / F

# Passage 4 Inscriptions

GROUP 4

True (T) / False (F)

Scan me with ChemFord

1. This geometry is known as octahedral — T / F

2. The equatorial groups are separated by a bond angle of 90º... — T / F

3. $SF_6$ adopts this geometry and is symmetrical, hence exhibiting no net dipole moment... — T / F

4. Molecules adopting this geometry exhibit Berry pseudorotation... — T / F

5. The axial groups are separated by a bond angle of 180º... — T / F

6. This geometry has a steric number of six... — T / F

7. This geometry can have up to 8 electron groups, hence the prefix 'octa'... — T / F

8. Replacing the axial bonding groups of this geometry with lone pairs gives rise to a square planar geometry... — T / F

9. $ClF_3$ and $MnCl_5^{2-}$ both adopt this geometry... — T / F

**"I should write my analysis of the inscriptions here... This should allow me to work out which passage to take and where the untrustworthy students are...If an inscription is false, I should state why..."**

Passage 1 Inscriptions:

Passage 2 Inscriptions:

Passage 3 Inscriptions:

Passage 4 Inscriptions:

Which passage should we take? ________

Extra: Which group has the untrustworthy students? ________

*The narrative contains a key clue for this (audio log V-04)!*

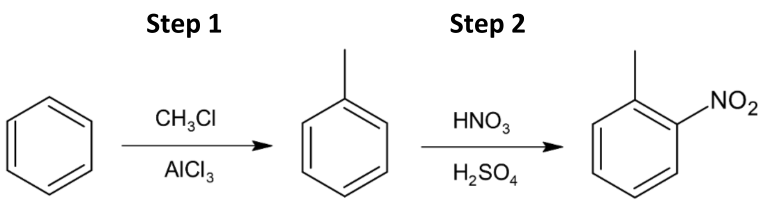
# Appendix H – Activity Measurement Rubric

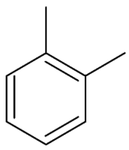
| Score given (per inscription) | | | Explanation (Student writes about) | |
|---|---|---|---|---|
| 2 | Same | Group 1 | Inscription 1 | $H_2O$ and $SCl_2$ are both bent/angular |
| | | | Inscription 2 | Geometry is known as bent/angular |
| | | | Inscription 3 | $BeCl_2$ is linear |
| | | | Inscription 4 | Lone pairs repel more strongly than bonding pairs |
| | | | Inscription 5 | Geometry has a steric number of 3 or 4 depending on the number of lone pairs present |
| | | | Inscription 6 | Double bonds contribute one bonding group |
| | | | Inscription 7 | Bond angles in bent/angular are less than those in linear geometries |
| | | | Inscription 8 | Adding an additional bonding group would result in trigonal planar/trigonal pyramidal |
| 1 | More or less | | Student response does not include what is stated above/insufficient explanation | |
| 0 | No answer | | Blank/no evidence of reasoning about the question | |
| 2 | Same | Group 2 | Inscription 1 | Equatorial groups are separated by an angle of 120° |
| | | | Inscription 2 | Equatorial and axial groups are separated by a bond angle of 90° |
| | | | Inscription 3 | The equatorial and axial groups are not equivalent |
| | | | Inscription 4 | Trigonal bipyramidal has 5 bonding groups |
| | | | Inscription 5 | This geometry exhibits Berry pseudorotation |
| | | | Inscription 6 | Removing a bonding group would result in the seesaw geometry |
| | | | Inscription 7 | The two axial groups are separated by a bond angle of 180° |
| | | | Inscription 8 | Geometry is called trigonal bipyramidal |
| | | | Inscription 9 | $PF_5$ adopts this geometry |
| 1 | More or less | | Student response does not include what is stated above/insufficient explanation | |
| 0 | No answer | | Blank/no evidence of reasoning about the question | |
| 2 | Same | Group 3 | Inscription 1 | $CH_4$ and $CF_4$ both adopt this geometry |
| | | | Inscription 2 | Bond angles are separated by 109.5° |

| | | | | |
|---|---|---|---|---|
| | | | Inscription 3 | Symmetrical molecules are non-polar |
| | | | Inscription 4 | Square planar and seesaw geometries both have 4 bonding groups |
| | | | Inscription 5 | This geometry has a steric number of four and no lone pairs |
| | | | Inscription 6 | No Berry pseudorotation is observed in tetrahedral molecules |
| | | | Inscription 7 | The name of this geometry is tetrahedral |
| | | | Inscription 8 | $CH_3Cl$ adopts this geometry and displays a dipole moment |
| | | | Inscription 9 | Phosphate ion is tetrahedral |
| | | | Inscription 10 | Bonding groups are each located at the corner of a tetrahedron |
| 1 | More or less | | Student response does not include what is stated above/insufficient explanation | |
| 0 | No answer | | Blank/no evidence of reasoning about the question | |
| 2 | Same | Group 4 | Inscription 1 | This geometry is known as octahedral |
| | | | Inscription 2 | Equatorial groups are separated by a bond angle of 90° |
| | | | Inscription 3 | $SF_6$ adopts an octahedral geometry and is symmetrical |
| | | | Inscription 4 | Octahedral molecules do not exhibit Berry Pseudorotation |
| | | | Inscription 5 | Axial groups are separated by 180° |
| | | | Inscription 6 | Octahedral molecules have a steric number of 6 |
| | | | Inscription 7 | Octahedral molecules can only have up to 6 electron groups |
| | | | Inscription 8 | Replacing axial groups with lone pairs gives rise to square planar geometries |
| | | | Inscription 9 | $ClF_3$ is T-shaped and $MnCl_5^{2-}$ is square bipyramidal |
| 1 | More or less | | Student response does not include what is stated above/insufficient explanation | |
| 0 | No answer | | Blank/no evidence of reasoning about the question | |

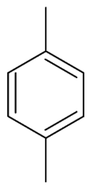
# Appendix I – $S_E$Ar test instrument

1. A student is producing nitrotoluene from benzene via the following reaction. The student observes that large quantities of compounds (i) and (ii) are produced during step 1. How can the amounts of (i) and (ii) be minimised?



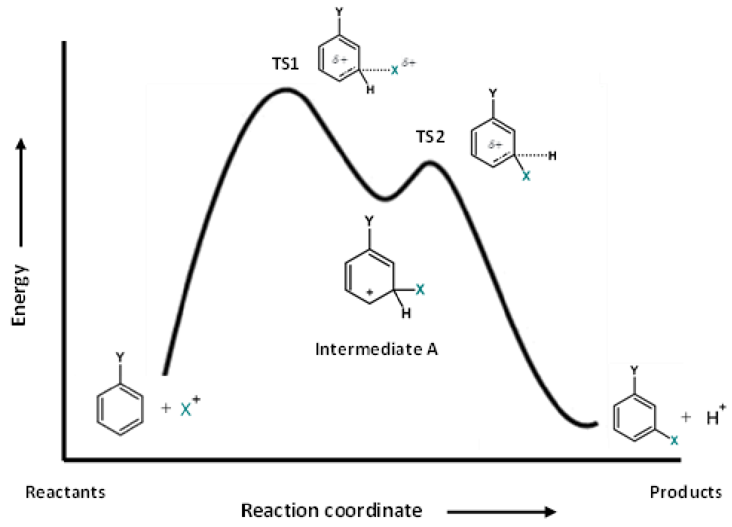
**(i)**



**(ii)**

   a. Use a reduced amount of the $AlCl_3$ catalyst with the same amount of $CH_3Cl$
   b. Carry out the reaction in step 1 at a lower temperature
   c. Use an excess amount of benzene relative to $CH_3Cl$
   d. Use a 1:1 ratio of benzene and $CH_3Cl$, with $FeCl_3$ as the Lewis acid catalyst

2. Which of the following statements is **correct** with regards to the potential energy diagram for the following electrophilic aromatic substitution reaction?

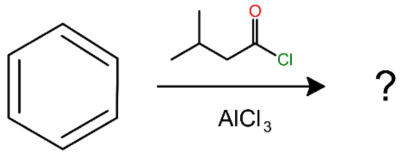



   a. Intermediate A does not contain a $sp^3$ hybridised carbon atom
   b. Substituent Y is likely to be an electron withdrawing group
   c. TS2 represents the rate-limiting transition state
   d. The addition of a deactivating substituent on the ring structure would decrease the activation energy.

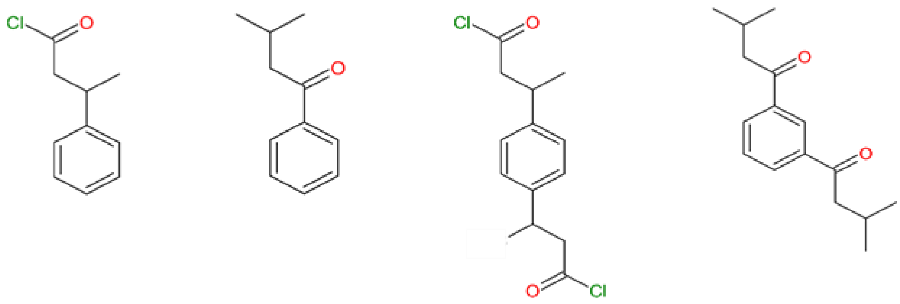
3. Which of the following statements is **correct**?

   a. An aromatic compound retains its aromaticity in the first step of an electrophilic aromatic substitution reaction
   b. The presence of a methyl group ring substituent destabilises the arenium ion intermediate
   c. The final step of an electrophilic aromatic substitution reaction is the gain of a proton
   d. The rate determining step of an electrophilic substitution reaction is the formation of the intermediate sigma complex

4. Which of the following is the **major** product for the following reaction?



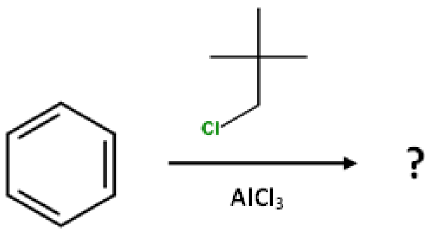

1 2 3 4

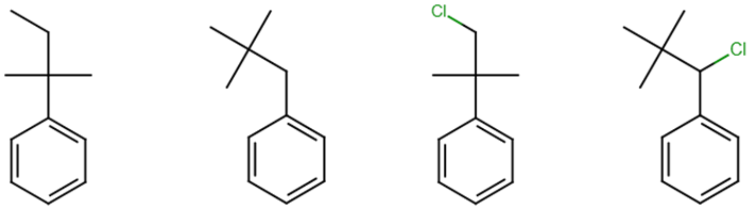



a. 1 c. 3

b. 2 d. 4

5. Which of the following is the **major** product for the following reaction?




1 2 3 4

a. 1 c. 3

b. 2 d. 4

6. In an electrophilic aromatic substitution reaction, amino groups are ortho/para directors because:

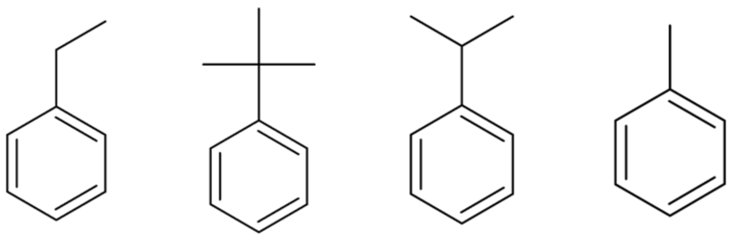
    a. The amino group stabilises the intermediate carbocation by an inductive effect. Charge distribution is greatest at the ortho and para positions.

    b. The amino group stabilises the intermediate carbocation by a resonance effect. Charge distribution is greatest at the ortho and para positions.

    c. The amino group destabilises the intermediate carbocation by a resonance effect. Charge distribution is lowest at the meta position.

    d. The amino group destabilises the intermediate carbocation by an induction effect. Charge distribution is lowest at the meta position.

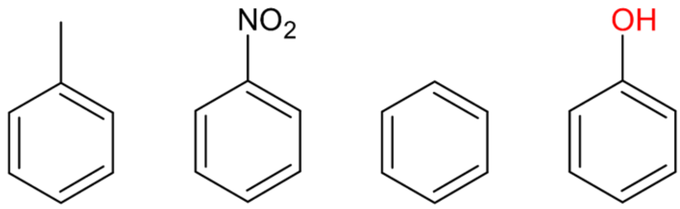
7. Four alkylbenzenes are shown below. Which of these compounds would you expect to produce the most **para** isomer from a sulfonation reaction?



**1** **2** **3** **4**

    a. 1
    b. 2
    c. 3
    d. 4

8. Rank the following compounds in order of **decreasing** reactivity towards electrophilic aromatic substitution.




1 2 3 4

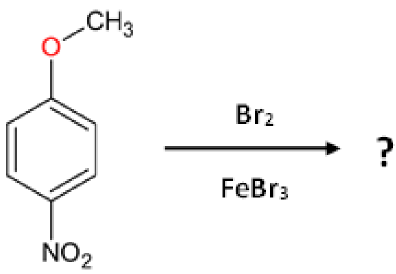
a. 2 > 4 > 1 > 3

b. 1 > 3 > 2 > 4

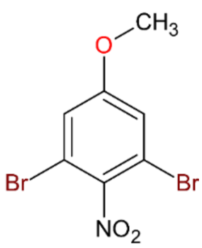
c. 4 > 1 > 3 > 2

d. 3 > 2 > 4 > 1

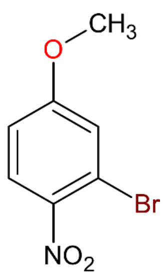
9. Which of the following is the **major** product of the following reaction?



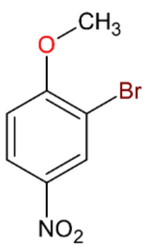

a.




b.




c.




d. Products (a), (b), and (c) are formed in equal amounts.

# Appendix J – Interview schedule for worked examples activity

*[Present student with the teaching material used in the session]*

**Question A1**
What did you like about the way the material was presented to you?

**Question A2**
What did you dislike about the way the material was presented to you?

**Question A3**
What would you like to add or remove from the particular material?

*[Concept discussion with student on the knowledge of the subject. Show examples to student]*

**Question B1**
What electrophilic aromatic substitution reactions do you know? Can you provide examples?
*If Friedel-Crafts is mentioned, move to question C1*

**Question B2**
What are the stages of an electrophilic aromatic substitution reaction?
*Refer to the two-step process of the reaction if the student is having trouble.*

**Question B3**
What are ortho/para and meta directing groups? Could you provide examples?
*If activating and deactivating groups come up, move to question D1*

**Question B4**
What is the rate-determining step in an electrophilic aromatic substitution reaction?

**Question C1**
Can you name any key differences between the Friedel-Crafts alkylation and acylation reactions?
*If rearrangement comes up, ask about carbocation stability.*
*If product is more/less nucleophilic comes up, ask about subsequent alkylation and how this may be minimised experimentally*
*If activating and deactivating groups come up, move to question D1*

**Question C2 & C3**
In the following example, which of the molecules:

a) Will act as the nucleophile
b) Will become the electrophile
c) Will act as the Lewis acid catalyst

**Question D1**
What is the difference between an activating and deactivating group? Could you provide examples?

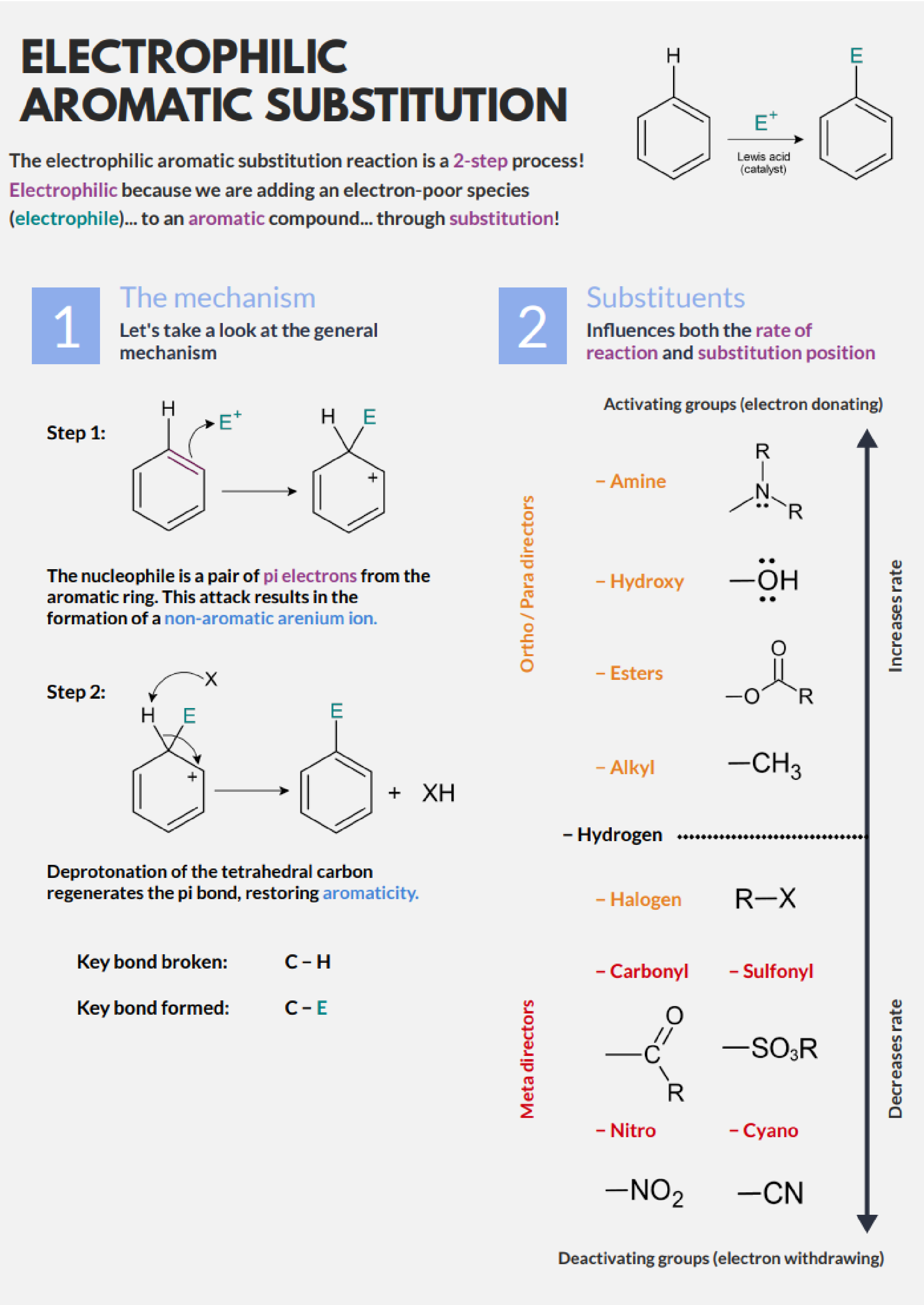
*If rate of reactivity is not mentioned, ask how each group may impact the rate of reaction.*

**Question D2**
If you have two substituents on a ring, which one would control the substitution position and why?
*Show example if student is struggling.*

**Question D3**
How does the intermediate stability affect the position of substitution?
*Electron-withdrawing = meta (likely), electron donating = ortho/para etc.*

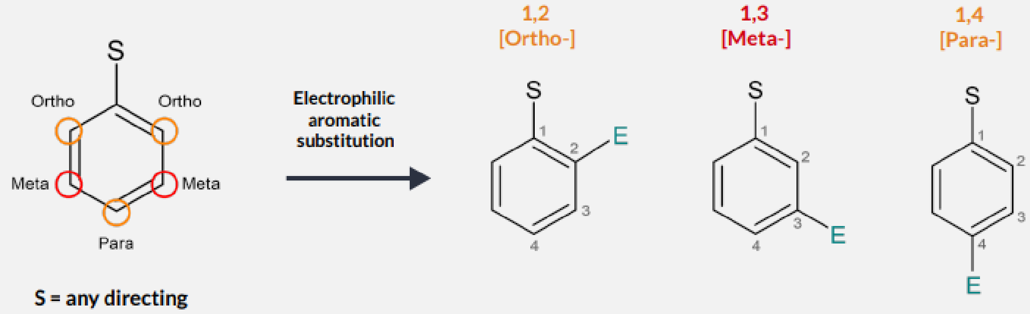
**Question D4**
What is the difference between resonance stabilisation and inductive stabilisation?

# Appendix K – Worked examples activity



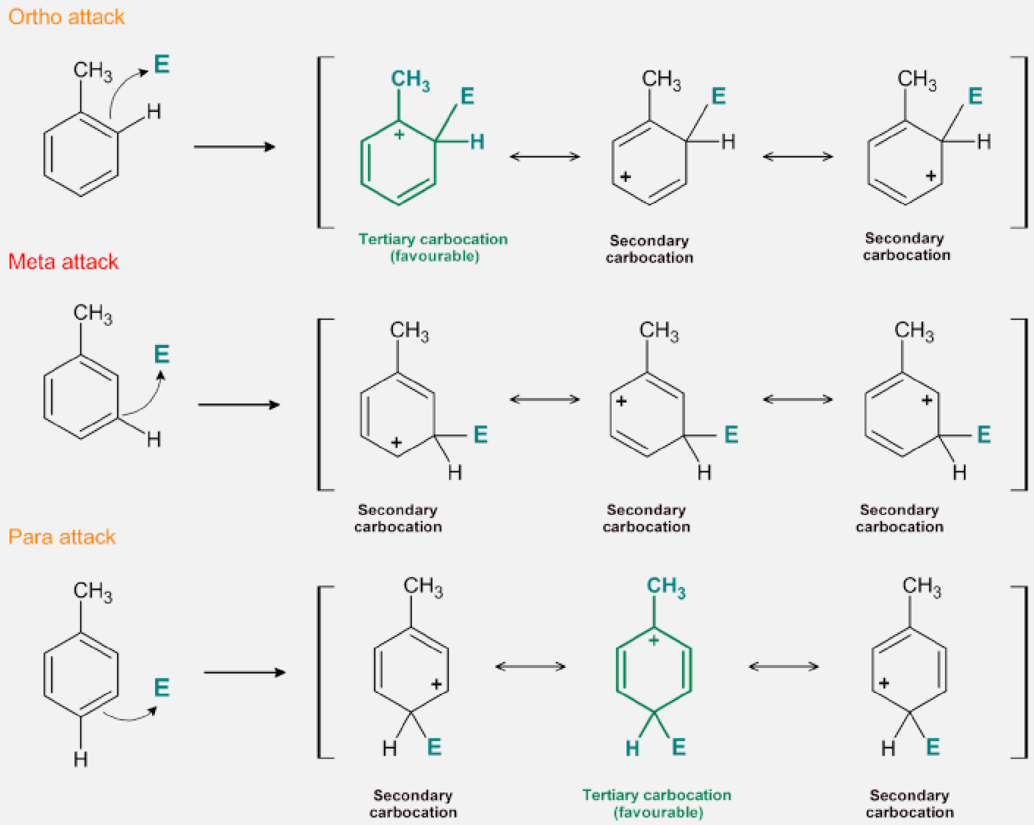
# 3 Revisiting terms

**Let's revisit three key terms: Ortho, Meta and Para positions**




**S = any directing substituent**

**The stability of the intermediate influences the regioselectivity! If we look at the electrophilic aromatic substitution of toluene, which has an ortho/para directing methyl group.**



## 4 Key reactions

**We've summarised 6 key electrophilic aromatic substitution reactions:**

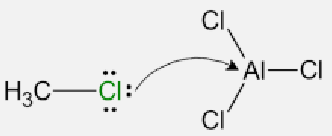
| Name | Reagent | Lewis acid | Electrophile |
|---|---|---|---|
| Nitration | $HNO_3$ | $H_2SO_4$ | $N^+O_2$ |
| Sulfonation | $SO_3$ | $H_2SO_4$ | $S^+O_3H$ |
| Friedel-Crafts Alkylation | R-X | $AlCl_3/FeCl_3$ | $R^+$ |
| Friedel-Crafts Acylation | R-COX | $AlCl_3/FeCl_3$ | $R-C^+=O$ |
| Bromination | $Br_2$ | $AlBr_3/FeBr_3$ | $Br^+$ |
| Chlorination | $Cl_2$ | $AlCl_3/FeCl_3$ | $Cl^+$ |

## 5 A completed example (using ChemFord)

**Let's look at the detailed mechanism of a Friedel-Crafts Alkylation on toluene. Our haloalkane reagent (R–X) will be chloromethane.**

→ **Let's import both aluminium trichloride and chloromethane from the app inventory!**
**(To do this, toggle from "scan" to "play" mode).**

**The electrophile (methyl) is generated, as well as aluminium tetrachloride. Let's move these aside for now!**




**STEP 1**
→ **Import toluene from the app inventory.**
**As the methyl electrophile molecule is moved closer to the toluene molecule, nucleophilic attack occurs from the pi electrons on the aromatic toluene ring. This attack results in a loss of aromaticity of the ring. This is the rate determining step!**

**STEP 2**
→ **A resonance stabilised non-aromatic intermediate (arenium ion) is generated from the attack on the electrophile. As methyl groups are ortho/para directing, substitution occurs at two positions: (i) ortho; and (ii) para.**

**The resonance button can be used to toggle between resonance forms. Which carbocations are secondary and tertiary?**

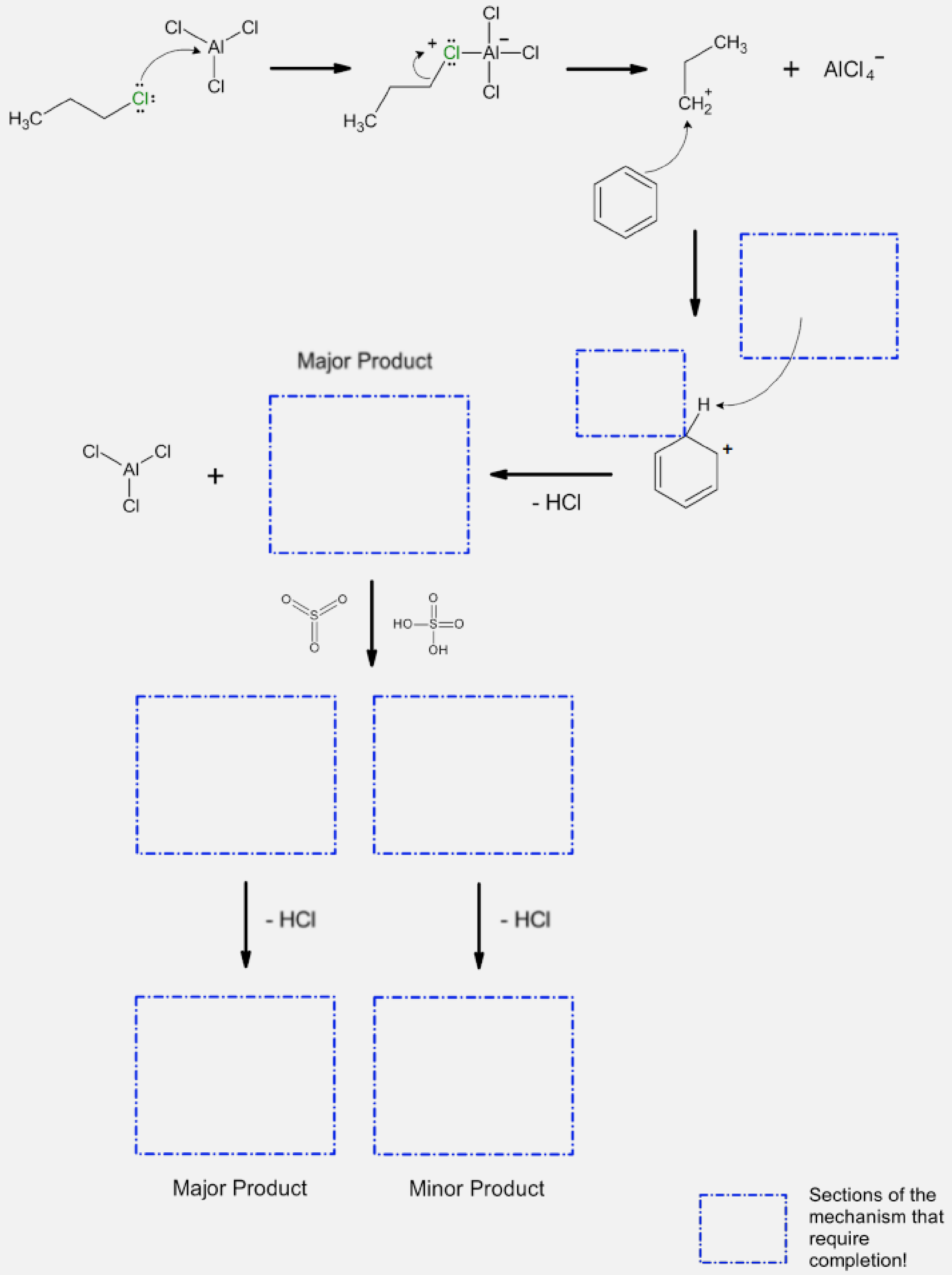
**STEP 3**
→ **Moving the tetrahedral aluminium tetrachloride close to the arenium ion intermediate results in loss of a proton. This restores the aromaticity of the final product and generates HCl as a side product.**

**The xylene product is more nucleophilic than the toluene reagent so further alkylation can take place. This is unique to the Friedel-Craft Alkylation.**

**Important: The carbocation electrophile will rearrange to form a more stable secondary/tertiary structure when bonding if possible. This is unique to the Friedel-Craft Alkylation.**

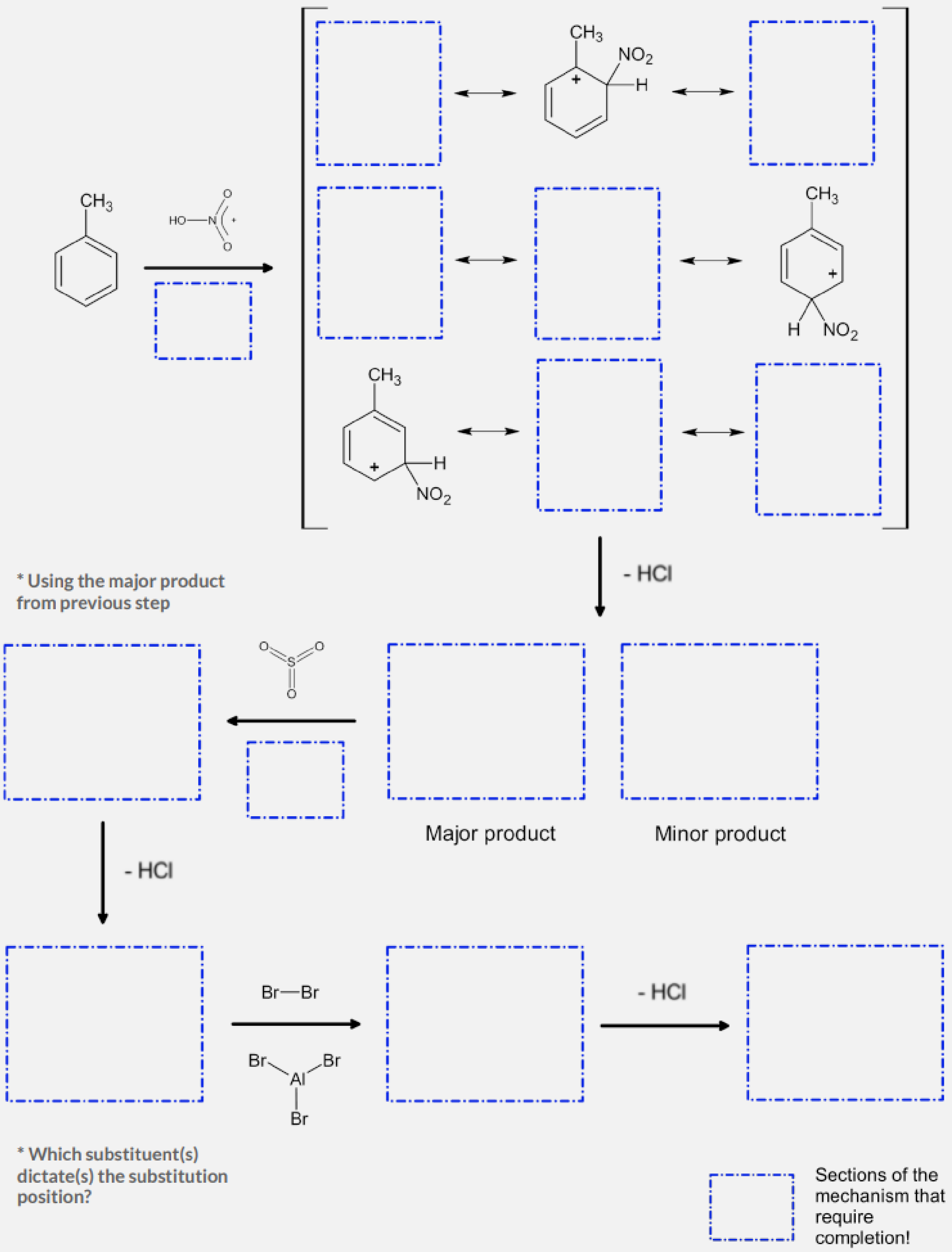
## 6 Worked example 1

**Now that we've seen an example of an electrophilic aromatic substitution reaction, apply these concepts to finish the partial mechanism below!**



## 7 Worked example 2

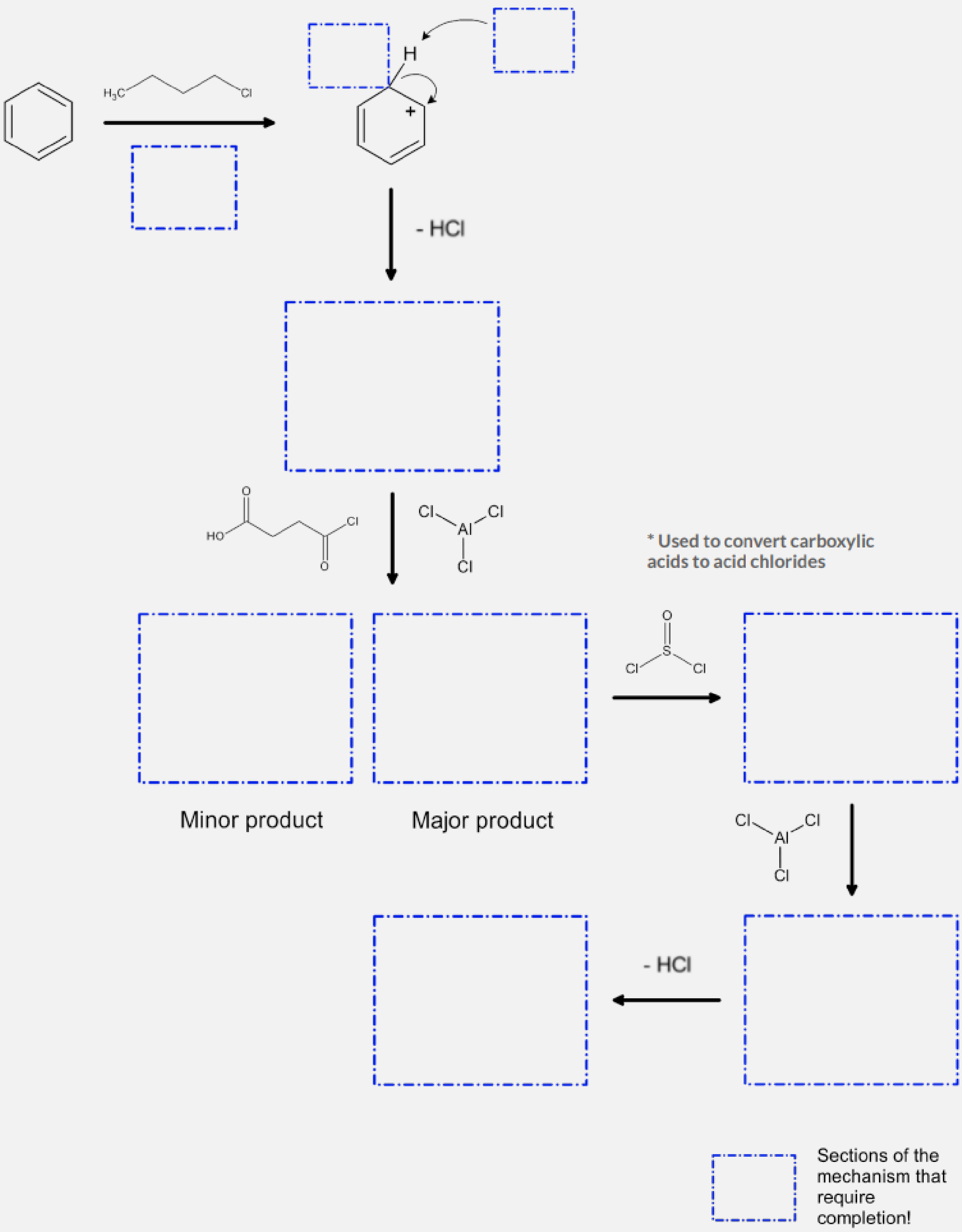
**Congratulations on completing the previous example! Have a go at the next example.**



8

## Worked example 3

**You're almost there! Complete the last example to prove to yourself that you have mastered the concepts!!**



# Appendix L – Negotiated codebook for worked example activity

| Code | Code Description |
|---|---|
| Positive student feelings | Any positive statements regarding student satisfaction and session facilitation |
| Negative student feelings | Any negative statements regarding student satisfaction and session facilitation |
| Future developments/fixes | Any reference to future designs/improvements for AR technology and/or booklet. |
| Booklet content and presentation | Any student statements regarding design aspects/content within the learning material (booklet) |
| Cognition | Student references to areas of working memory/visualisation |
| Reflection on the learning process | Any statement where students reflect on their own learning |
| Positive views on material design | Any statements regarding positive material design. |
| Negative views on material design | Any statements regarding negative material design. |
| Use of augmented reality | Any statements regarding views/usage of the augmented reality tool |
| **ELECTROPHILIC AROMATIC SUBSTITUTION DISCUSSION** | |
| Activating and deactivating groups | Any statements regarding activating and deactivating groups demonstrating either evidence of understanding or misunderstanding. |
| Ortho, meta, and para groups | Any statements regarding ortho, meta, and para groups demonstrating either evidence of understanding or misunderstanding. |
| 'Which substituent controls?' | Any statements regarding addition of substituent(s) when the aromatic ring has existing groups: demonstrating either evidence of understanding or misunderstanding. |
| Knowledge of types of $S_EAr$ reaction | Any statements where students can/cannot provide examples of electrophilic aromatic substitution reactions when prompted. |
| Reaction Example 1 | Any discussion points around the topic of reaction example 1 |
| Reaction Example 2 | Any discussion points around the topic of reaction example 2 |
| Friedel-Crafts (FC) (rearrangement) | Any statements around the topic of rearrangement in the FC alkylation reaction |

| | |
|---|---|
| Friedel-Crafts (FC) (subsequent reactivity) | Any statements around the topic of unwanted subsequent reactivity in the FC alkylation reaction |
| Reaction mechanism – rate determining step (RDS) | Any statements around the topic of the RDS in the $S_EAr$ mechanism |
| Reaction mechanism – role of Lewis acid | Any statements around the role of the Lewis acid in the $S_EAr$ mechanism |
| Reaction mechanism – regioselectivity | Any statements around regioselectivity in the $S_EAr$ mechanism |
| Reaction mechanism – aromaticity | Any statements around intermediate/product aromaticity in the $S_EAr$ mechanism |