Preprints.org

Article

Using Virtual Reality to Learn Abstract Chemistry Concepts: An Experimental Study Using a VR Chemistry Lab and Molecule Simulation

Elliot Hu-Au [*]

Article

Using Virtual Reality to Learn Abstract Chemistry Concepts: An Experimental Study Using a VR Chemistry Lab and Molecule Simulation

Elliot Hu-Au

Montclair State University; huaue@montclair.edu

Abstract: In this 2 x 2 between-subjects experimental study, a virtual reality (VR) laboratory simulation is paired with a VR molecular world intervention to teach chemistry concepts. The independent variables are the implementation timing of the molecular world intervention (Pre-lab vs. Integrated) and the level of embodiment (Traditional vs. VR). Eighty students (N = 80), ages 11-18 years old, from a community center in New York City participated, completing a pretest, the laboratory simulation, the molecular intervention, and a post-test. The pre- and post-test measures included multiple-choice, free-response, and drawing questions. A key finding was that integrating the intervention within the lab simulation, no matter which level of embodiment, led to significantly higher gains in learning. The combination of using physical manipulatives and integrating them within the lab exercise (Integrated Traditional condition) demonstrated the greatest gains overall. On drawing measures, the Integrated VR condition showed significant improvement in three out of the four drawing categories (i.e., molecule shape, atom quantity, and relative sizes). The implications are that even though using a VR molecular world intervention can lead to significant learning of abstract chemistry content, the use of physical manipulatives is still a more effective tool.

Keywords: virtual reality; educational virtual reality; educational technology; computer-assisted instruction; educational activities; chemistry

1. Introduction

Chemistry is a notoriously difficult subject to learn for many primary and secondary-level students. The combination of abstract concepts, specialized vocabulary with nuanced meanings, unique mathematical explanations, and few obvious connections with daily life often confuses novice chemistry students (Johnstone, 2007; Broman & Simon, 2015; Berg, Orraryd, Pettersson, & Hultèn, 2019). While chemistry experts already have surmounted these obstacles, new students require guidance to see the intricacies of the discipline. Effective teaching methods can be vital to both the accessibility of the subject matter but also in maintaining the interest of the students. Unfortunately, emphasis on new teaching strategies is often neglected in chemistry classes, where traditional teaching methods are prolific and are also considered "boring" and "uninteresting" to many students (Broman & Simon, 2015, p. 1260).

This research study examines the potential of virtual reality (VR) as a tool for learning chemistry. Its novelty can be highly motivational for students and its affordances for visuospatial learning are promising (Dalgarno & Lee, 2010). This study utilizes a laboratory simulation that has been observed to provide a similar experience to a real-life laboratory (Hu-Au & Okita, 2020) and a molecular world intervention that uses either physical manipulatives or a VR experience. A pre-posttest design is utilized, including assessments with written and drawing modalities. Analysis of different categories of chemistry knowledge is also conducted.

2. Background

Chemistry is a critical field when considering science, technology, engineering, and mathematics (STEM) subjects. The difficulties that students encounter when learning chemistry can be long-lasting and far-reaching. They can often lead to discouragement and eventual withdrawal from not just chemistry, but related fields in STEM (Halim, Rahman, Wahab, Mohtar, 2018; Broman & Simon, 2015). This is due to many of chemistry's topics being foundational for understanding other disciplines such as biology and physics (Herrman-Abell, Koppal, & Roseman, 2016; Özmen & Ayas, 2003). Difficulties learning chemistry also have the added consequence of disproportionately deterring underrepresented minoritized students and women from pursuing STEM degrees and careers (National Science Foundation, 2014). If chemistry education can be improved it can have a ripple effect in improving many other areas of STEM education.

Promising chemistry teaching strategies center on using better pedagogy to teach abstract concepts within the subject. Many of these effective pedagogies are founded upon Alex Johnstone's (1982) framework of the triplet nature of chemistry knowledge. He viewed the topics in chemistry as split into three different categories: the macro, symbolic, and sub-micro. The macro (or macroscopic) level knowledge refers to aspects of chemistry that can be seen and measured empirically, the symbolic level being abstract and containing the representations using equations and notation, and the sub-micro level being the abstract mental models and images of the molecular world. Johnstone (1982) proposed that emphasizing the relationships between these three levels of knowledge would increase the clarity with which students could understand the chemical world. Since his initial explanation of the triplet relationship of chemistry knowledge, many researchers have observed greater learning by students when the macro-symbolic-sub-micro relationships are directly addressed (Berg, Orraryd, Pettersson, & Hulten, 2019; Jaber & BouJaoude, 2012; Barnea and Dori, 1999; Johnstone, 2007).

2.1. The Chemistry Laboratory as a Connector of Macro and Sub-Micro Knowledge

The chemistry laboratory, or practical work, is an experience where students gain hands-on experiences conducting chemistry experiments. It is an experience where the combination of macro and sub-micro knowledge is presented in many ways but also where students have many difficulties (Berg et al., 2019; Tsaparlis, 2009). Laboratory work is primarily an activity centered on macro-level knowledge, where students have hands-on interactions with substances, observe laws, and measure changes. Tsaparlis (2009) speculates that the level of stimuli and complication of laboratory experiments can potentially overload a student's capacity to process information, keeping them from properly analyzing the experiment and its underlying mechanisms in the sub-micro realm. Tan, Goh, Chia, and Treagust (2009) have observed similar behaviors where students blindly follow laboratory instructions without thinking of the sub-micro level. Others have found that novices generally think at the macro level when confronted with new information and need to be prompted to think at the sub-micro level (Al-Balushi, 2012). It is highly possible that the experiences in the laboratory may be easily observed at the macro level but never even considered at the sub-micro level. It also could be argued that considering the sub-micro realm is not even a natural instinct for many novice chemistry students. These factors contribute to a high chance of missing the important relationship between chemistry's macro and sub-micro levels, undermining the potential of the laboratory experience.

2.2. Virtual Reality as a Tool for Learning Abstract Chemistry Knowledge

To address this macro-sub-micro weakness in chemistry education, an intervention was developed using virtual reality (VR). VR is a tool that has the potential to address this weakness due to its ability to simulate immersive, macro-level chemistry experiences while also enabling interactions with sub-micro models. These VR lab simulations are capable of enabling similar levels of learning when compared to real-life lab experiences (Hu-Au & Okita, 2020). Regarding learning at the sub-micro level, VR poses an interesting use case where immersive digital environments can be used to study these abstract concepts. Previous research shows VR to be an effective tool at visualizing and simplifying complex abstract concepts (Lamb & Etopio, 2020; Ferrell et al., 2019; Merchant et al., 2013; Dalgarno & Lee, 2010). In addition, conducting a laboratory experiment in VR

3

allows one to control the timing of processes and bring attention to sub-micro mechanisms at appropriate times, an ability that is rarely available in real-life laboratories.

These previously mentioned affordances of VR technology demonstrate a promising tool for exploring the macro-sub-micro relationship in chemistry. Although this has been widely recognized, research studies on VR use in chemistry are not that frequent. Previous studies involving VR and chemistry have centered on simulations of outside environments (Fung et al., 2019), replicated laboratory environments for training purposes (Tatli & Ayas, 2013), or just polled stakeholders in the field without actually conducting an experiment (Suleman et al., 2019). In one similar research study at a higher education institution, Ferrell et al. (2019) tested an immersive, microscopic VR world with an organic chemistry class. They showed that the intervention led to an increase in learning but did not mention if the control group was given a comparable learning activity. So, it is possible the students learned more in VR because they simply were able to have more time exposed to the content. Thus, this study seeks to fill this gap in knowledge surrounding pedagogy and VR's efficacy to assist students in learning abstract chemistry concepts.

2.3. Learning Theory Framework

The possible differences between conditions are framed by certain aspects of different theories of cognition and learning. Differences in learning results based on the level of embodiment could provide evidence for its effects in understanding abstract chemistry content. For instance, "immersion" is a frequently cited affordance of VR that has little clarity in learning theory. It most closely resembles Csikszentmihalyi's flow theory (1990), in that the user is so immersed in an experience that they are in a state of extreme concentration and oblivious to any outside stimuli. Supporting research proposes that a high level of immersion allows the students' mind to focus solely on the content, which may lead to greater understanding (Minocha, Tudor, & Tilling, 2017).

Barsalou's grounded cognition theory (2008) also provides the conceptual framework for the physical gestures and environment involved in our VR intervention and its potential to clarify specific abstract concepts. For instance, having the user move their hands towards each other to connect the digital representations of molecules should reinforce the idea of atoms combining to create a molecule. Furthermore, positioning the user as a microscopic object in a molecular world environment should create a bodily "simulation" (p. 618). By doing this, a user gains a multimodal experience that can reinforce the abstract concept they are focused on. Utilizing physical manipulatives does this as well but to a lesser extent. The varying levels of embodied movements could demonstrate nuances between these levels and their link to creating knowledge. The use of physical manipulatives, with fine motor motions of finger gestures are much more intricate but less expansive and fit into what Johnson-Glenberg & Megowan-Romanowicz (2017) call the 2nd level of "embodied education taxonomy" (p.2). The use of a VR headset and controllers, with high levels of sensorimotor engagement, gestural congruency, and immersion fits into the 4th level, meaning that there is a difference in embodiment that can be used to analyze learning results.

Finally, the importance of visuospatial cues for learning abstract knowledge (Tversky, 2005) supports our efforts in the VR simulation to reify atoms, molecules, and their unseen behaviors. Our utilization of different colors, shapes, and sizes for atoms are spatial representations that ease the user's mental load for recognizing reactive properties (a common strategy among chemical experts) (Hegarty & Stull, 2012). The VR modality allows designers to have greater flexibility over the appearance of artifacts, allowing designers to match colors, scale shapes more accurately, and even scale weights to more clearly signify specific characteristics. Physical manipulatives are more limited, being uniform in size, weight, and randomly colored, restricting their efficacy in representing multiple aspects of atomic content.

2.4. Research Questions

This study seeks to explore these research questions:

1. How does the timing of a VR sub-micro experience in a laboratory exercise affect a student's ability to learn sub-micro and macroscopic chemistry knowledge?

2. How does the level of embodiment in a VR sub-micro experience affect a student's ability to learn sub-micro and macroscopic chemistry knowledge?

3. Methodology

This 2 x 2 between-subjects study examined the learning results from manipulating the chronological order and level of embodiment of a VR chemistry intervention and laboratory exercise. Quantitative methodology, utilizing statistical analysis on pre- and post-test measures was used to determine if there were significant differences in learning gains between conditions. Each condition involved an interactive VR laboratory simulation and either a session of playing with a set of traditional molecular physical manipulatives (i.e., ball and stick models) or interacting with an immersive VR molecule modeling program.

3.1. Research Sample

The data collection of the study, Teachers College IRB #20-434, took place from March - November 2022. The research sample was 80 students, ages 11-18 years old, randomly selected from a local community center in a multicultural, socio-economically diverse area of New York City. The study was conducted during the after-school and summer camp sessions. 96% of the students were of African-American, Hispanic, or of both descents. Gender balance was 42 females and 38 males. Over 80% of the students were between 11-14 years old. Rationale for conducting the study outside of schooltime was to minimize lost instructional time due to the study participation. Participants completed the study in a single session, taking from 80-95 minutes. Students were compensated with a $10 Amazon gift card at the completion of their session. Most students had already been introduced to the basic concepts of atoms and molecules in their schools. About half of the students had experienced VR in some form prior to this study. More detailed demographic statistics for the sample population are represented in Appendix A.

3.2. Study Design & Procedure

There are four conditions into which students were randomly divided, 20 students per condition (see Figure 1). The control condition was using physical manipulatives as the intervention as a pre-lab exercise (Pre-Lab Traditional). Condition 2 used the VR intervention as the pre-lab exercise (Pre-Lab VR). Condition 3 used the physical manipulatives at a specific time point during the lab exercise (Integrated Traditional). Condition 4 used the VR intervention at a specific time point during the lab exercise (Integrated VR).

The procedure of the research study had commonalities between the conditions but differed in the timing and modality of the intervention. Each condition began with the demographic survey, pretest, and viewing of the Lab Safety Video. After this, each condition diverged with the Integrated and Pre-lab VR conditions having VR practice time while the control condition started with the physical manipulatives intervention. The Pre-lab conditions conducted the intervention and Manipulatives Assessment prior to conducting the laboratory exercise. The Integrated conditions conducted the laboratory exercise and participated in the intervention immediately after the chemical reaction began. The detailed procedure for each condition is detailed in Figure 2.

			Timing Placement of Sub-micro experience	
			Pre-Lab	**Integrated**
Embodiment Level: Mode of manipulati		**Traditional** (imaginative experience with physical manipulatives**)**	1. Physical manipulatives experience *before* the lab	3. Physical manipulatives experience *within* the lab

	Virtual Reality (immersive embodied experience with virtual manipulatives)	2. VR experience *before* the lab	4. VR experience *within* the lab

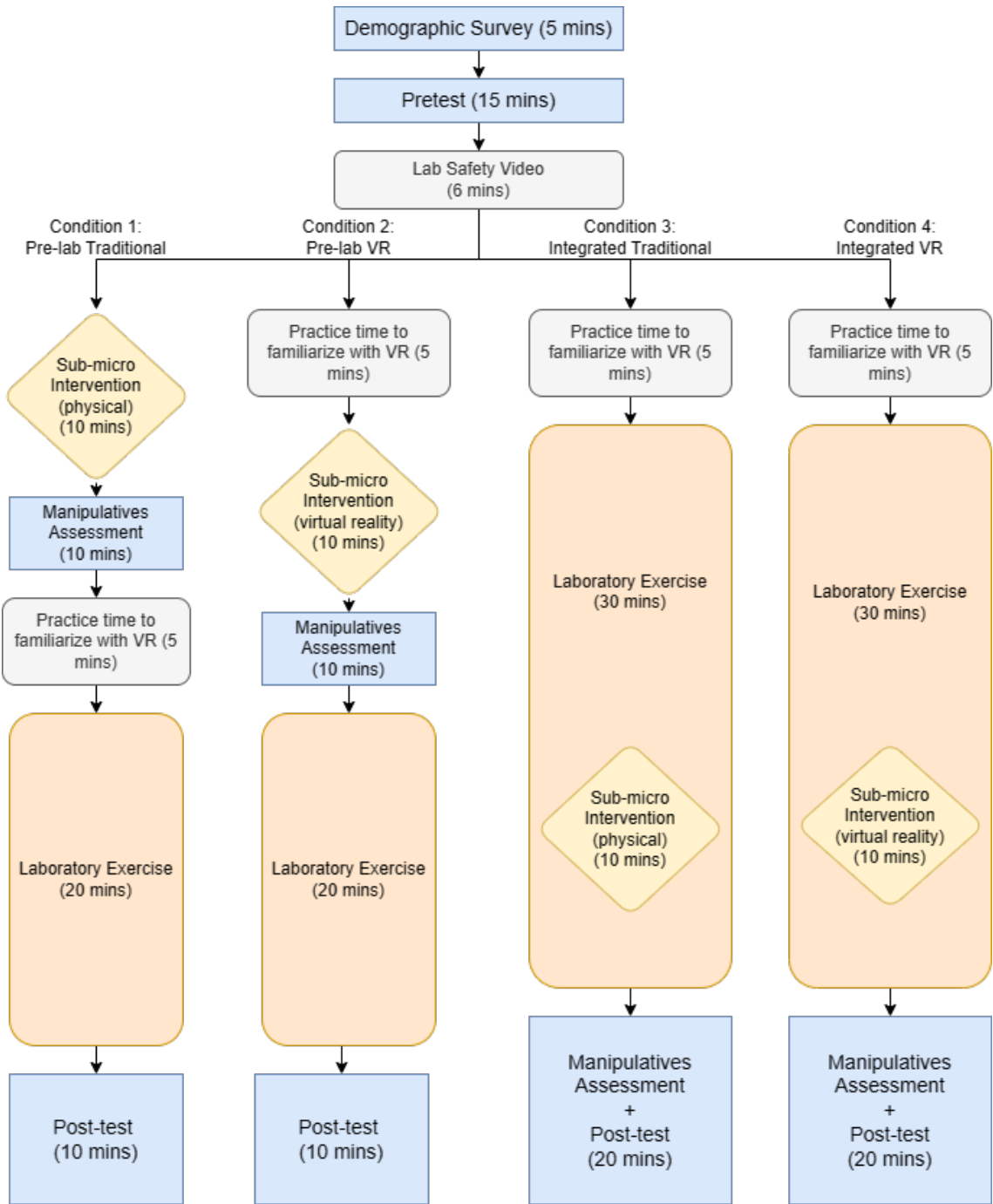**Figure 1.** The 2 x 2 Between-Subjects Design for this Study.

Figure 2. Flowchart of the Study Procedure.

3.3. The VR Chemistry Lab Environment and Exercise

The VR laboratory exercise is centered on the Next Generation Science Standard MS-PS1-2: Matter and its Interactions (NGSS Lead States, 2013). The laboratory exercise involves mixing anhydrous copper (II) chloride with water then introducing a piece of aluminum metal to the aqueous

solution (See Figures 3-5). The aluminum then combines with the chlorine atoms, replacing the copper in a single-replacement reaction. This lab experiment was chosen because it is recommended for middle to high school chemistry students by the American Association of Chemistry Teachers (AACT, 2020). Furthermore, it is a reaction that can be comprehended by younger students as a simple demonstration of compounds breaking apart and other compounds forming. Older students can observe the reaction on deeper levels, examining ionic compounds and balancing equations. The VR hardware used in this study were the *Oculus Quest* and *Meta Quest 2* (Oculus, 2020).
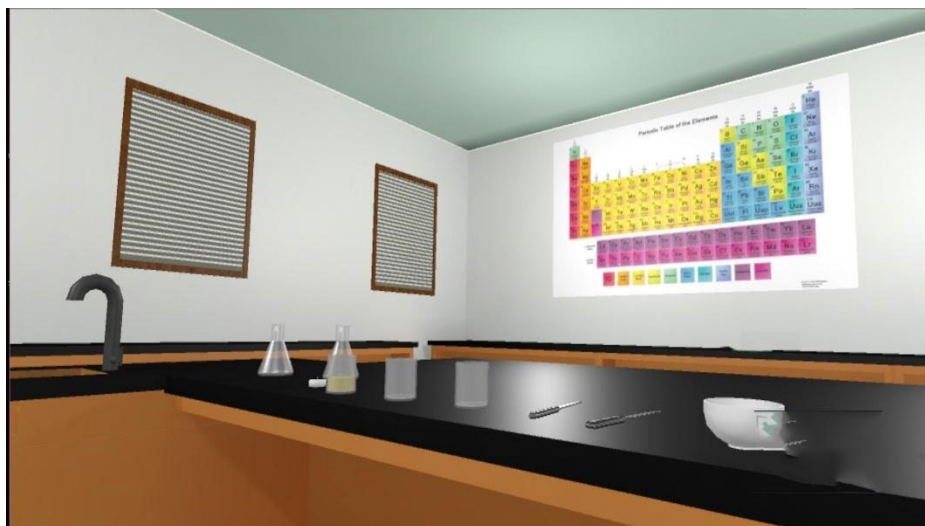

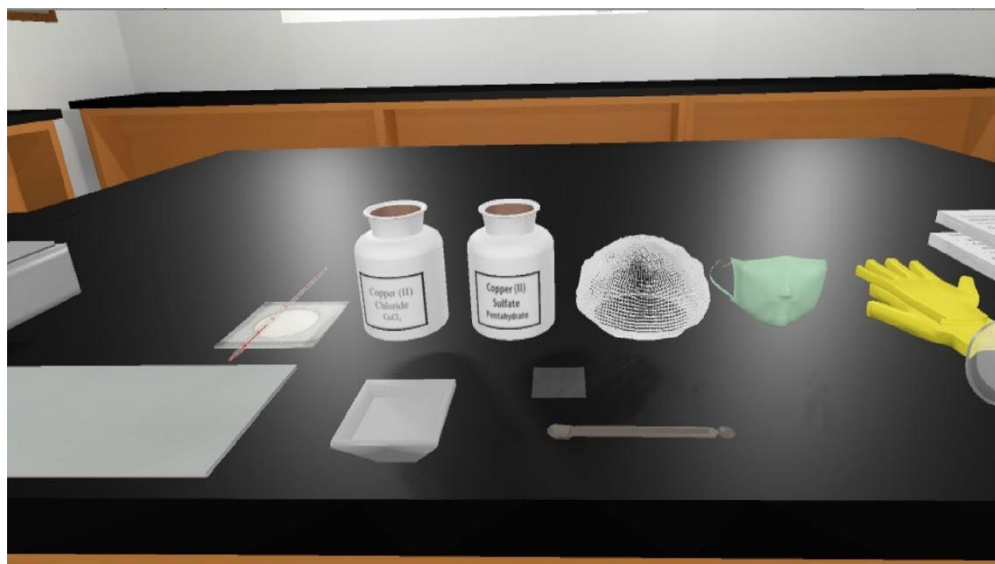**Figure 3.** The Digital Environment of the VR Chemistry Lab Simulation.

Figure 4. The Experiment Materials used in the VR Lab Simulation.

Figure 5. Smoke and Audio Sounds Simulating Reactions in VR.

3.4. The Sub-Micro Intervention

3.4.1. Traditional Version (Physical Manipulatives)

The Physical Manipulatives version of the sub-micro intervention was an introductory lesson on atoms, molecules, and atomic behavior. It utilized a "ball-and-stick" molecule set, physically manipulable objects that are used traditionally in chemistry classes (see Figure 6). The lesson that incorporated these manipulatives was designed by the primary investigator and verified as appropriate by two subject matter experts.
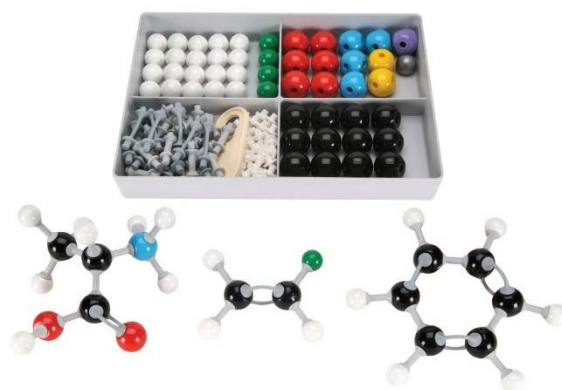

Figure 6. Traditional "Ball-and-Stick" Chemistry Molecule Set used in the Physical Manipulatives Lesson.

3.4.2. Virtual Reality Version

The Virtual Manipulatives Lesson was a VR experience designed and developed by the principal investigator (Figure 7 & 8). It utilizes the advantages of VR to immerse the user in the "molecular world" where atoms and molecules are at the same scale as the participant. The experience takes place inside of a beaker, as if the user is witnessing the chemical reaction at the molecular level. Within the lesson, the user is guided verbally by an outside experimenter. Almost all of the procedure

of this intervention mimics what was done in the traditional physical manipulatives version. The script used is identical to the Physical Manipulatives script except for a few references to the controls in VR. Both lessons are aligned with the underlying chemical reaction present in the macroscopic lab exercise.
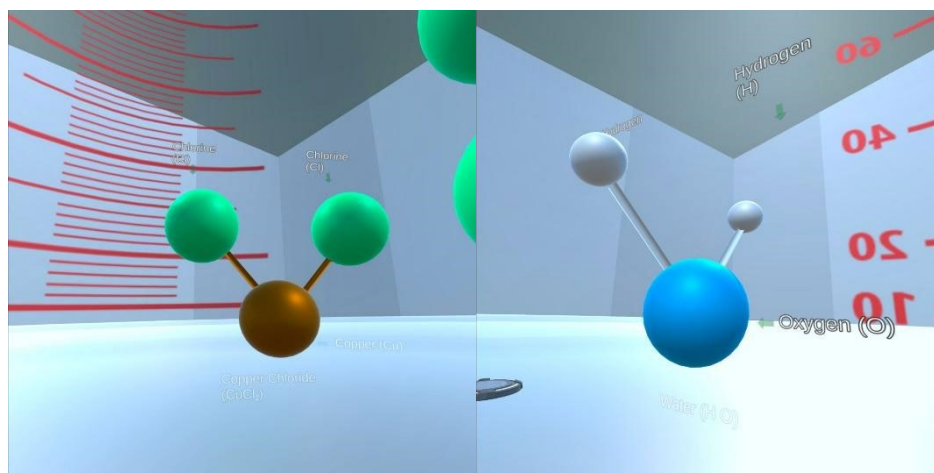

Figure 7. Screenshots from the VR Manipulatives Intervention showing models of the $CuCl_2$ (left) and H_2O molecules (right) prior to mixing.
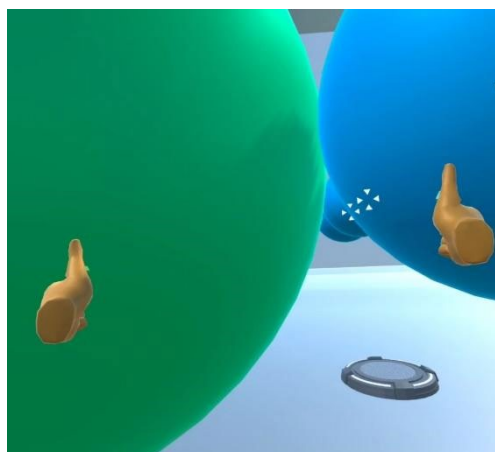

Figure 8. Screenshot from the VR Manipulatives Intervention demonstrating how the user can virtually "grasp" atoms and bring them together to bond them to each other.

3.5. Data Collection Measures

Internal reliability of the measures was assessed by looking at their Cronbach's alpha value. This checks the consistency of the measurement instruments to ensure all questions are accurately measuring the same concept. SPSS software was used to compute the Cronbach's Alpha for the pretest, manipulative assessment, and post-test. Their values were all greatly above the "good acceptance" value of $\alpha > 0.7$ (Ursachi, Horodnic, & Zait, 2015) except for the "acceptable" value for the manipulative assessment ($\alpha = .656$).

Since the pretest questions all had parallel questions on the post-test, reliability for the measure was similar to the item difficulty distribution as seen for the post-test (see Appendix B). Development and validation of this instrument, as well as the manipulative assessment and post-test, followed Adams and Wieman's (2010) framework for instrument design. Questions were developed by the principal investigator in collaboration with a subject-matter expert. Each question was constructed to align to Johnstone's (1982) categories of chemistry knowledge: sub-micro, macroscopic, and symbolic. In this manner, analysis could illuminate any resulting growth in each category.

Each measure was administered on a laptop using Qualtrics software. Some questions required drawing, so a pencil and paper were also provided. A short description of each is below.

3.5.1. Pretest

The pretest was conducted at the beginning of the session. The students individually completed the 35-question pretest: 19 multiple-choice (MC), 12 free-response (FR), and 4 drawing (D) questions. Examples of a few of the questions are below:

MULTIPLE CHOICE (Sub-micro): The bond between hydrogen and oxygen in water is strong compared to the bond between sodium and chlorine in salt. Which molecule's bonds would take more energy to break?

A) Water   B) Salt   C) The same amount of energy is needed for either

FREE RESPONSE (Sub-micro): What is the difference between an atom and a molecule?

3.5.2. Manipulatives Assessment

The manipulative assessment was designed to measure any learning gains strictly from the pre-lab versions of the sub-micro intervention. It was intended to show whether the VR version of the sub-micro intervention has any benefits over the traditional, physical manipulatives version without any interference by the VR lab simulation. This assessment had 13 questions (6 MC, 6 FR, 1 D) that pertain to the concepts covered in the manipulatives lesson. The item difficulty distribution for this measure is shown in Appendix B as well.

MULTIPLE CHOICE (Macroscopic): What did you observe happen to the temperature of the mixture when the copper chloride was placed into the water?

A) It increased   B) It stayed the same   C) It decreased

FREE-RESPONSE (Symbolic): What should be the formula for the newly created molecule involved in this reaction?

$Al + Cl + Cl + Cl \rightarrow$ _____

3.5.3. Post-Test

The post-test had 22 questions (13 MC, 6 FR, 3 D) that covered similar concepts to the pretest, but some questions utilized information specific to the laboratory experiment. For the Integrated conditions, the Manipulatives Assessment was combined with the post-test to be one summative assessment. This was designed to be able to match the content of every pretest question.

3.5.4. Drawing Measures

Two questions on each of the pre- and post-tests required the participants to draw a "ball-and-stick" model of certain molecules. Special attention was given to these questions as they represented students' attempts at visualizing abstract content. The grading rubric was adapted from a subject matter expert's publicly available "molecular model construction" rubric (StudyLib, 2022). Drawings were scored based on the accuracy of the elements used, number of specific atoms, relative angle between atoms, and the relative size between atoms.

Two drawing questions were also designed to measure the students' ability to connect macroscopic chemistry information to sub-micro knowledge. These questions utilized an image of some object (e.g., solid pieces of metal floating in a flask of clear liquid, cup of orange juice) and asked students to draw what the molecules of objects in the picture might look like and how they are arranged. A successful drawing would demonstrate an understanding of individual particles, their physical proximity to each other, and their position or structure in relation to other particles. An example of a drawing question and two students' responses is below (see Figure 9).

DRAWING Q.: A carbon atom will combine with four hydrogen atoms and create the stable molecule of methane, CH_4. If you had the objects below (see Figure 6) how would you create a model of this molecule? Draw a picture of it. Use balls to represent atoms and the sticks to be bonds. Label the names of the atoms and where the bonds are.
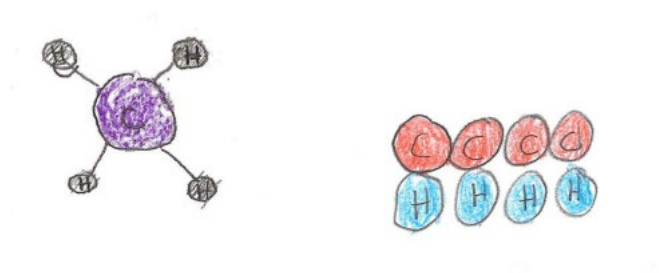
Figure 9. Samples of two student answers to the above drawing question.

3.6. Data Analysis and Synthesis

3.6.1. Pretest – Post-Test Measures

An inter-rater reliability (IRR) process was conducted for the free-response and drawing questions (Hallgren, 2012); five cases were randomly chosen and five raters independently scored all them. The scores were compared, a Krippendorff α value computed (Hayes & Krippendorff, 2007), and if the $\alpha < 0.80$, any discrepancies were discussed and the common rubric was edited. Then five new cases would be chosen and the process repeated until the value of $\alpha > 0.80$. Once this satisfactory value was achieved, the remaining cases were divided equally amongst the raters and individually evaluated.

Paired samples *t*-tests were used on each condition to analyze the score differences between the pretest and post-test. Questions in each measure were categorized into topics of symbolic, sub-micro, and macro and compared from Pretest to Post-test. For the 2x2 conditions, a two-way ANOVA was used to determine if there were any main and interaction effects due to the two independent variables. If a statistically significant difference between the groups was found, then a Tukey post hoc test was used to determine which groups were significantly different.

4. Results

4.1. Overall Score Results for Pre-Lab vs Integrated Conditions

A paired samples *t*-test was conducted to determine if any significant changes occurred within the symbolic, sub-micro, or macroscopic categories from the pretest scores to the post-test. The two opposing conditions were the *Pre-Lab* and *Integrated* conditions. Results showed both conditions scored significantly better on the post-test in Symbolic and Sub-micro categories (*Pre-Lab* Symbolic: M = 14.6%, t = 3.820, p <.001; Sub-micro: M = 7.6%, t = 2.617, p = .013; *Integrated* Symbolic: M = 16.4%, t = 4.344, p <.001; Sub-micro: M = 13.7%, t = 4.515, p <.001). However, both conditions declined significantly in the Macroscopic category (*Pre-Lab*: M =-19.4%, t = -6.76, p <.001; *Integrated*: M = -10.6%, t = -4.227, p <.001) (see Table 1).

Table 1. Paired Samples t-Test for SSM Categories for Timing Conditions.

Condition	Pretest			Posttest			ΔM	t(39)	p	Cohen's d
	M	*SD*	*SE*	*M*	*SD*	*SE*				
1. Pre-lab (control)										
Symbolic	**31.3**	**22.4**	3.5	46.0	24.8	3.9	14.7	3.820	<.001*	.604
Sub-micro	31.7	23.1	3.7	39.4	16.7	2.6	7.6	2.617	.013*	.414
Macroscopic	43.3	19.9	3.1	23.8	12.8	2.0	-19.4	-6.757	<.001*	-1.068
2. Integrated										
Symbolic	26.9	21.3	3.4	43.3	26.6	4.2	16.4	4.344	<.001*	.687
Sub-micro	28.9	22.2	3.5	42.5	15.2	2.4	13.7	4.515	<.001*	.714
Macroscopic	39.7	19.9	3.1	29.1	13.5	2.1	-10.6	-4.227	<.001*	-.668

* p < .05

4.2. Overall Score Changes for Embodiment Level Conditions

A paired samples t-test was conducted to determine if there were any significant changes in symbolic, sub-micro, and macroscopic scores based on embodiment level. Results showed that the Traditional condition significantly improved in the symbolic (ΔM = 24.8%, t = 7.358, p <.001) and sub-micro categories (ΔM = 13.9%, t = 4.242, p <.001) but decreased in macroscopic (ΔM = -15.3%, t = -5.048, p <.001). The VR condition significantly improved in sub-micro (ΔM = 7.44%, t = 2.828, p = .007) but decreased in macroscopic knowledge (ΔM = -14.7%, t = -5.839, p <.001) (see Appendix C).

4.3. Manipulative Assessment – Comparison of Embodiment Levels

A paired samples t-test showed that neither condition changed scores significantly from pretest to the manipulative assessment. A one-way ANOVA determined that the differences between conditions was not significant (F(1)=.595, p = .445, η^2 = .015).

4.4. Overall Score Changes for 2x2 Conditions

Data analysis was also conducted to compare each of the four conditions within the 2x2 testing matrix. The goal was to determine if there was any interaction between the variables that significantly affected their scores from pretest to post-test. Overall score changes, scores on the chemistry triplet of knowledge categories, molecule drawings, and macro-sub-micro drawings were analyzed for significant differences between the four conditions.

A paired samples *t*-test was conducted on each condition to determine whether they changed significantly from the pretest to the post-test. It was found that only condition 3 improved significantly (ΔM = 8.35%, SD = 15.3%, SE = 3.4%, t = 2.438, p = .025) (see Table 2).

Table 2. Paired Samples t-Test for Overall Score Changes for 2x2 Conditions.

	Condition	Pretest			Posttest			Diff	t(19)	p	Cohen's d
		M	SD	SE	M	SD	SE				
1.	**Traditional Pre-lab (control)**	**36.5**	**19.7**	**4.4**	38.4	13.9	3.1	1.94	.656	.520	.147
2.	Virtual Reality Pre-lab	34.2	18.5	4.1	29.6	13.7	3.1	-4.60	-1.674	.111	-.374
3.	Traditional Integrated	34.5	17.9	4.0	42.8	11.3	2.5	8.35	2.438	.025*	.545
4.	Virtual Reality Integrated	29.3	19.7	4.4	31.8	15.3	3.4	2.45	1.108	.282	.248

* p < .05

A two-way ANOVA was then conducted to determine if any of these conditions or the interactions between conditions had significant differences between each other. There was no statistically significant effect of the interaction between the timing variable and the embodiment variable (F(1) = .012, p = .913). However, a significant difference was identified within the timing conditions (F(1) = 5.513, p = .021) and the embodiment conditions (F(1) = 4.701, p = .033). No post hoc test was necessary since each condition had only two levels. The Integrated condition increased significantly more than the Pre-Lab condition and the Traditional condition increased significantly more than the VR condition (See Appendix D).

4.5. Molecule Drawings Analysis – Scores on Specific Categories for 2x2 Conditions

A two-way ANOVA was used to search for significant differences between conditions and their interaction for each specific drawing criteria. There was a significant interaction revealed between

the Timing and Embodiment variables for the Atom Quantity category (F(1) = 5.808, p = .018). There were also significant main effects for Embodiment conditions for the Names of Elements (F(1) = 6.774, p = .011) and the Relative Sizes of Atoms (F(1) = 4.057, p = .048) (see Table 3). The Tukey HSD test for pairwise comparisons was then used to distinguish which conditions were actually significantly different (see Appendix E). There was a significant difference between the Traditional Integrated and the VR Integrated conditions in three categories. Traditional was significantly greater in recalling Element Names (ΔM = .550, p = .046). Whereas, VR was significantly greater in recalling Atom Quantity (ΔM = .600, p = .035) and Relative Sizes of Atoms (ΔM = .650, p = .010).

Table 3. ANOVA Results for Molecule Drawing Category Scores for 2x2 Conditions.

Source	Sum of Squares	df	Mean Square	F (1)	p	η^2
Timing Condition						
Element Names	.800	1	.800	1.084	.301	.014
Atom Quantity	.313	1	.313	.317	.575	.004
Molecular Shape	2.113	1	2.113	2.142	.147	.027
Relative Sizes of Atoms	1.80	1	1.800	2.980	.088	.038
Embodiment Condition						
Element Names	5.00	1	5.00	6.774	.011*	.082
Atom Quantity	.313	1	.313	.402	.528	.005
Molecular Shape	.313	1	.313	.317	.575	.004
Relative Sizes of Atoms	2.45	1	2.45	4.057	.048*	.051
Interaction						
Element Names	.050	1	.050	.068	.795	.001
Atom Quantity	4.513	1	4.513	5.808	.018*	.071
Molecular Shape	.313	1	.313	.317	.575	.004
Relative Sizes of Atoms	1.80	1	1.80	2.980	.088	.038
* p < .05						

5. Discussion

5.1. Integrating a Sub-Micro Intervention within the Lab Exercise is Significantly Better for Learning

Research question 1 explored how changing the timing of intervention implementation would affect sub-micro knowledge gains. Both the pre-lab and integrated conditions showed significant growth in symbolic and sub-micro knowledge scores. This is a positive sign that a sub-micro-focused intervention either before or during a lab exercise will improve students' learning of abstract, sub-micro chemistry content.

The differences between the learning outcomes due to the timing conditions could be due to the lower amount of time between engaging with the intervention content and experiencing the related observable phenomena in the lab exercise. Traditionally, pre-labs are used in classroom laboratory settings to prime the students for learning specific topics and ease their cognitive load as they conduct the experiments (Agustian & Seery, 2017). While pre-labs have been proven to be helpful for learning (Winberg & Berg, 2007), there could be a recency effect at work here. This effect, where more recently observed objects or concepts are more easily retrieved from memory (Baddeley & Hitch, 1993), could explain why the Integrated condition scored better on the post-test.

Pedagogically, these results reinforce the idea that "just-in-time" learning could be a more effective learning method for laboratory experiences (Boese, 2016, p. 341). With "just-in-time" learning, a student can pause the situation they are in and quickly research the content they are confused or curious about. For a science lab, when students reach a point in the exercise where they encounter an exciting development and are curious about the underlying reasons, that is the best time for explicit information to be presented to the student (Killi & Morrison, 2015). Rather than rely on the student remembering an explanation that was encountered earlier or hoping that an instructor

is nearby, the timely explanation is now coupled with the instigating event and a stronger connection is made.

The ability to pause time and allow the student to delve deeper into the underlying scientific principles was one of the primary reasons for the implementation of the lab exercise in VR. Doing the lab exercise with real chemicals and equipment is valuable but would not afford the student the ability to pause the reaction and immediately seek out relevant information. The VR simulation allows explanations to be available exactly when the student encounters the most stimulating part of the reaction. They can also have the time to reflect and think about the hidden mechanisms behind what they have just observed. This advantage of VR, its ability to control time in an immersive context, can provide valuable instances of learning that would normally be inaccessible.

5.2. Traditional Hands-On Embodiment has Advantages over Virtual Reality Embodiment

Research question 2 focused on the effects of the level of embodiment for the participant during the intervention. It was hypothesized that participants would make better connections between sub-micro and macroscopic knowledge in the VR conditions due to the immersive model environment. However, this was not confirmed by the results. It is possible that the learning affordances of the physical manipulatives in the Traditional conditions are strong enough to overcome the affordances presented by immersive VR. As seen in other STEM education research, physical manipulatives are sometimes preferred by students over virtual ones and can lead to greater motivation towards learning (Justo et al., 2022).

Another explanation is that immersive VR is too distracting for students. This is consistent with findings from Makransky, Terkildsen, and Mayer (2019), who found that a highly immersive VR experience led to less fact learning for students than a low-immersion desktop experience. Their explanation was that the high level of immersion led to cognitive overload for the learner and distracted them from the material they were supposed to learn. For many students, especially those who were experiencing VR for the first time, being in an immersive VR experience was quite novel and distracting. Perhaps with more experience in VR they would be able to focus more on the tasks at hand.

It is possible that the physical manipulatives experience was a stronger grounding mechanism for cognition than the VR experience, leading to a better ability to extrapolate the information to other situations. The participants who used the physical manipulatives were able to feel the textures and physical weights of the objects representing atoms, whereas the VR participants may have been affected by the tactile gap between visually seeing certain objects, but physically only holding the hand controllers. From an embodied cognition perspective, this is not entirely surprising. Though VR enthusiasts champion embodied cognition as a supportive theory for comparable learning in VR, there must be a limit to its support when the learning tool is unable to fully replicate the sensorial experience. The current state of VR hand controllers is this limitation.

It is also possible that the disparity in presence of the experimenter in each condition could have contributed to the participants' academic outcomes. In the case of the Traditional condition, where the experimenter was in-person and sat near the participant, this level of human presence was more tangible than the VR condition, where the experimenter spoke instructions from outside the VR headset and did not have a digital avatar in the VR experience. A comparable situation is the different experiences of in-person learning vs. remote learning for students. Tomasik, Helbling, and Moser (2021) found that the majority of students ages 9-15 years old learned more than twice as fast in-person when compared to remote learning. This seems to be an age-specific effect, as studies comparing in-person and online learning for students over 18 years-old observed little difference in conceptual understandings between the two groups (Raes, 2022). Thus, an argument could be made against VR as a primary educational tool for young people or that designs of educational VR environments need to contain avatars or other teacher-like entities to assist young learners.

5.3. Benefits of VR for Abstract Sub-Micro Knowledge Learning

One important point from this comparison is that even though the Traditional condition showed greater learning than VR in some categories, both conditions showed significant improvement in the sub-micro knowledge realm on the post-test. Research supports this result, as students have been observed to remember more information by experiencing it in VR (Allcoat & von Mühlenen, 2018). The gestures in both modes of the intervention provided a concrete situation with which the abstract concept could then be associated (Kiefer & Pulvermuller, 2012). The VR intervention also led to significant improvements in identifying correct sizes and atom quantity of different molecules in the drawing measures, which were helpful assessments of the students' visuospatial abilities. The VR environment allowed digital objects to be rendered at various sizes, whereas the traditional ball-and-stick models were all manufactured to be equal sizes.

5.4. Limitations

5.4.1. Inadequate Sample Size

There are a few limitations that affect the breadth of the results from this study. First, the sample size was not at the preferred number to attain a large effect size (0.80) for significant results. A conservative pre-study power analysis found that to achieve a FWER significance level appropriate to the standard $\alpha = 0.05$, each condition would need N = 39 participants per condition. However, this was difficult to achieve due to the length of each individual study session.

5.4.2. Alignment of the Macro-Sub-Micro Connection in Measures

From the poor results in macroscopic knowledge of students in all conditions, it is clear that addressing macroscopic and sub-micro chemistry concepts together is difficult. One potential fault in this study may lie in the alignment of the measures with the design of the sub-micro intervention. The low values for item discrimination for some questions in the measures allude to this. For example, the intervention presented relatively little modeling of the inner atomic structure of solids. However, most of the questions connecting macroscopic and sub-micro content involved visualizing the inner structure of solid objects, which led to poor performance on the post-test measures by students. In addition, reminders of the macroscopic changes after the reaction were not robust enough in the sub-micro intervention. It was hoped that the return to the lab simulation for the *Integrated* conditions would implicitly create that connection. However, a more overt scaffold may be more effective since guided scaffolding activities are often necessary in situations where students need to reflect on the connections between these worlds (Keiner & Graulich, 2021). Effective methods of teaching the macro-sub-micro connection in chemistry continue to be a pedagogical research need.

6. Implications

The results of this study are meant to inform practitioners of the usefulness of VR in chemistry education. It is safe to say that VR holds a motivational aspect for most students that can provide a boost to their interest. While it can be debated if this is compelling enough to warrant the use of VR in formal education settings, the confirmation of physical manipulatives as an effective learning tool is heartening. As most teachers have easier access to physical manipulatives, VR use can continue to inspire as an extracurricular tool and allow the traditional items to remain as everyday learning objects.

Implications for educational technology research is that there was shown to be significant differences in learning chemistry content using different levels of the embodied education taxonomy (Johnson-Glenberg & Megowan-Romanowicz, 2017). Specifically, content about atoms, molecules, and chemical reactions was remembered better by students who engaged in using physical manipulatives (level 2) than those who used immersive VR (level 4). This is not the only factor that could have affected the learning results but it is interesting that in this case, a lower level of embodiment resulted in greater learning outcomes. More research is needed in this area.

Additionally, researchers designing learning experiences for VR must agree that it is difficult to do well. Due to the almost limitless abilities one can provide the user in VR, the lack of boundaries

can cause an instructional designer to lose focus. Innocuous details in the VR world can easily become distractions that keep the learner from focusing on the learning objective. For chemistry education, this study represents a step in the direction of effective design of VR learning experiences. Knowing effective learning designs can lead to better chemistry experiences for students and more students pursuing this important field.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to participant privacy reasons.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A Study Sample Demographic Information

Appendix B Item Difficulty Distribution for Pre/Post-test and Manipulative Assessment

Appendix C Paired Samples t-Test for SSM Categories for Embodiment Level Conditions

Appendix D ANOVA Results for Overall Score Change for 2x2 Conditions

Appendix E Pairwise Comparisons for ANOVA for Timing vs. Embodiment Interaction on Molecule Drawing Categories

References

1. AACT. (2020). Classroom Resources: Observing a chemical reaction. Retrieved from https://teachchemistry.org/classroom-resources/observing-a-chemical-reaction
2. Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. International Journal of Science Education, 33(9), 1289–1312. https://doi.org/10.1080/09500693.2010.512369
3. Agustian, H. Y., & Seery, M. K. (2017). Reasserting the role of pre-laboratory activities in chemistry education: a proposed framework for their design. Chem. Educ. Res. Pract, 518, 518. https://doi.org/10.1039/c7rp00140a
4. Akuma, F. V., & Callaghan, R. (2019). A systematic review characterizing and clarifying intrinsic teaching challenges linked to inquiry-based practical work. Journal of Research in Science Teaching, 56, 619–648. https://doi.org/https://doi.org/10.1002/tea.21516
5. Al-Balushi, S. M. (2013). The effect of different textual narrations on students' explanations at the submicroscopic level in chemistry. Eurasia Journal of Mathematics, Science and Technology Education, 9(1), 3–10. https://doi.org/10.12973/eurasia.2013.911a
6. Allcoat, D., & von Mühlenen, A. (2018). Learning in virtual reality: Effects on performance, emotion and engagement. Research in Learning Technology, 26(1063519), 1–13. https://doi.org/10.25304/rlt.v26.2140
7. Baddeley, A. D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? Memory & Cognition, 21(2), 146–155.
8. Barnea, N., & Dori, Y. J. (1999). High-School Chemistry Students' Performance and Gender Differences in a Computerized Molecular Modeling Learning Environment. Journal of Science Education and Technology, 8(4), 257–271.
9. Barsalou, L. W. (2008). Grounded Cognition. Annual Review of Psychology, 59(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639
10. Berg, A., Orraryd, D., Pettersson, A. J., & Hultén, M. (2019). Representational challenges in animated chemistry: self-generated animations as a means to encourage students' reflections on sub-micro processes in laboratory exercises. Chemistry Education Research and Practice, 20(4), 710–737. https://doi.org/10.1039/c8rp00288f
11. Boese, E. (2016). Just-In-Time Learning for the Just Google It Era. SIGCSE '16: Proceedings of the 47th ACM Technical Symposium on Computing Science Education, 341–345. https://doi.org/10.1145/2839509.2844583

12.   Broman, K., & Simon, S. (2015). Upper secondary school students' choice and their ideas on how to improve chemistry education. International Journal of Science and Mathematics Education, 13(6), 1255–1278. https://doi.org/10.1007/s10763-014-9550-0

13.   Csikszentmihalyi, M. (1990). Flow: the psychology of optimal experience (1st ed.). New York, NY: Harper & Row.

14.   Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? British Journal of Educational Technology, 41(1), 10–32. https://doi.org/10.1111/j.1467-8535.2009.01038.x

15.   Ferrell, J. B., Campbell, J. P., McCarthy, D. R., McKay, K. T., Hensinger, M., Srinivasan, R., Zhao, X., Wurthmann, A., Li, J., & Schneebeli, S. T. (2019). Chemical Exploration with Virtual Reality in Organic Teaching Laboratories. Journal of Chemical Education, 96(9), 1961–1966. https://doi.org/10.1021/acs.jchemed.9b00036

16.   Fung, F. M., Choo, W. Y., Ardisara, A., Zimmermann, C. D., Watts, S., Koscielniak, T., Blanc, E., Coumoul, X., & Dumke, R. (2019). Applying a Virtual Reality Platform in Environmental Chemistry Education to Conduct a Field Trip to an Overseas Site [Product-review]. Journal of Chemical Education, 96(2), 382–386. https://doi.org/10.1021/acs.jchemed.8b00728

17.   Hadero, H., & Yamat, R. (2023, January 8). CES 2023: Smelling, touching take center stage in metaverse. Yahoo News. Retrieved from https://www.yahoo.com/now/ces-2023-smelling-touching-center-150040084.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAMiJ6jftG8pWQVwVXkBuWDtgA36vhsaM999QeS2JC2piBo_CLVEiC53HCckpMYe1z823G_8jrNoS1X4Q-YzAyeU6zVyIdqfed-sdfIQXEbMQDWyQTZy9JCtK3wYzeOJzfLoSx9RRUMOYC7MyB8knqglH_DfWWFaeCTXa1OvGVO9l

18.   Halim, L., Rahman, N. A., Wahab, N., & Mohtar, L. E. (2018). Factors Influencing Interest in STEM Careers: An Exploratory Factor Analysis. Asia-Pacific Forum on Science Learning and Teaching, 19(2).

19.   Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. In Tutorials in Quantitative Methods for Psychology (Vol. 8).

20.   Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. Communication Methods and Measures, 1(1), 77–89.

21.   Hegarty, M., & Stull, A. T. (2012). Visuospatial thinking. In K. J. Holyoak & R. G. Morrison (Eds.), Oxford library of psychology. The Oxford handbook of thinking and reasoning (pp. 606–630). Oxford University Press. https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199734689.001.0001/oxfordhb-9780199734689-e-31

22.   Herrmann-Abell, C. F., Koppal, M., & Roseman, J. E. (2016). Toward High School Biology: Helping Middle School Students Understand Chemical Reactions and Conservation of Mass in Nonliving and Living Systems | Enhanced Reader. CBE-Life Sciences Education, 15(74), 1–21.

23.   Hu-Au, E. & Okita, S. (2020, June 19-23). Exploring Differences in Student Learning and Behavior between Real-Life and Virtual Reality Chemistry Laboratory Experiments. The International Conference of the Learning Sciences, Nashville, TN. https://icls2020.org/

24.   Jaber, L. Z., & BouJaoude, S. (2012). A Macro-Micro-Symbolic Teaching to Promote Relational Understanding of Chemical Reactions. International Journal of Science Education, 34(7), 73–998. https://doi.org/10.1080/09500693.2011.569959

25.   Johnson-Glenberg, M. C., & Megowan-Romanowicz, C. (2017). Embodied science and mixed reality : How gesture and motion capture affect physics education. Cognitive Research: Principles and Implications, 2(1), 24. https://doi.org/10.1186/s41235-017-0060-9

26.   Johnstone, A. H. (1982). Macro- and micro-chemistry. School Science Review, 64, 377–379.

27.   Johnstone, A. H. (2007). Science Education: We know the answers, let's look at the problems. Proceedings of the 5th Greek Conference "Science Education and New Technologies in Education," 1–11.

28.   Justo, E., Delgado, A., Llorente-Cejudo, C., Aguilar, R., & Cabero-Almenara, J. (2022). The effectiveness of physical and virtual manipulatives on learning and motivation in structural engineering. Journal of Engineering Education, 111(4), 813–851. https://doi.org/10.1002/jee.20482

29.   Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. Cortex, 48(7), 805–825. https://doi.org/10.1016/J.CORTEX.2011.04.006

30.   Keiner, L., & Graulich, N. (2021). Beyond the beaker: Students' use of a scaffold to connect observations with the particle level in the organic chemistry laboratory. Chemistry Education Research and Practice, 22(1), 146–163. https://doi.org/10.1039/d0rp00206b

31.   Killi, S., & Morrison, A. (2015). Just-in-time Teaching, Just-in-need Learning: Designing towards Optimized Pedagogical Outcomes. Universal Journal of Educational Research, 3(10), 742–750. https://doi.org/10.13189/ujer.2015.031013

32.   Krathwohl, D. R. (2002). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives (Abridged Edition). Theory Into Practice, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

33. Lamb, R., & Etopio, E. A. (2020). Virtual Reality: a Tool for Preservice Science Teachers to Put Theory into Practice. Journal of Science Education and Technology, 29, 573–585. https://doi.org/10.1007/s10956-020-09837-5

34. Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learning and Instruction, 60(December 2017), 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

35. Merchant, Z., Goetz, E. T., Keeney-Kennicutt, W., Cifuentes, L., Kwok, O., & Davis, T. J. (2013). Exploring 3-D virtual reality technology for spatial ability and chemistry achievement. Journal of Computer Assisted Learning, 29(6), 579–590. https://doi.org/10.1111/jcal.12018

36. Minocha, S., Tudor, A. D., & Tilling, S. (2017). Affordances of Mobile Virtual Reality and their Role in Learning and Teaching. HCI 2017: Digital Make Believe - Proceedings of the 31st International BCS Human Computer Interaction Conference, HCI 2017, 2017-July. https://doi.org/10.14236/EWIC/HCI2017.44

37. National Science Foundation, (2014). STEM education data and trends 2014. Retrieved from https://nsf.gov/nsb/sei/edTool/data/college-11.html

38. NGSS Lead States. 2013. Next Generation Science Standards: For States, By States. Washington, DC: The National Academies Press.

39. Oculus. (2020). Oculus Main Website. Retrieved from www.oculus.com

40. Özmen, H., & Ayas, A. (2003). Students' difficulties in understanding of the conservation of matter in open and closed-system chemical reactions. Chemistry Education: Research and Practice, 4(3), 279–290.

41. Raes, A. (2022). Exploring Student and Teacher Experiences in Hybrid Learning Environments: Does Presence Matter? Postdigital Science and Education, 4, 138–159. https://doi.org/10.1007/s42438-021-00274-0

42. StudyLib. (2022 8 23). Molecular Model Construction Rubric. Referenced at: https://studylib.net/doc/5845804/molecular-model-construction-rubric

43. Suleman, M., Sugiyarto, H., & Ikhsan, J. (2019). Development of Media Three-dimensional (3D) Visualization using Virtual Reality on Chemistry Education. Journal of Physics: Conference Series. 1397, 12034. https://doi.org/10.1088/1742-6596/1397/1/012034

44. Tan., K. C. D., Goh, N. K., Chia, L. S., & Treagust, D. F. (2009). Linking the Macroscopic, Sub-microscopic and Symbolic Levels: The Case of Inorganic Qualitative Analysis. In J. K. Gilbert & D. Treagust (Eds.), Multiple Representations in Chemical Education (pp. 137–150). https://doi.org/10.1080/09500690903211393.

45. Tatli, Z., & Ayas, A. (2013). Effect of a virtual chemistry laboratory on students' achievement. Educational Technology & Society, 16(1), 159–170. http://www.ifets.info/journals/16_1/14.pdf

46. Tomasik, M. J., Helbling, L. A., & Moser, U. (2021). Educational gains of in-person vs. distance learning in primary and secondary schools: A natural experiment during the COVID-19 pandemic school closures in Switzerland. International Journal of Psychology, 56(4), 566. https://doi.org/10.1002/IJOP.12728

47. Tsaparlis, G. (2009). Learning at the macro level: the role of practical work. In J. K. Gilbert & D. Treagust (Eds.), Multiple Representations in Chemical Education (pp. 117–136). https://doi.org/10.1080/09500690903211393.

48. Tversky, B. (2005). Visuospatial Reasoning. In H. Pashler (Ed.), Cambridge Handbook of Thinking and Reasoning (pp. 209–240). https://doi.org/10.4135/9781452257044.n284

49. Ursachi, G., Horodnic, I. A., & Zait, A. (2015). How reliable are measurement scales? External factors with indirect influence on reliability estimators. Procedia Economics and Finance, 20, 679 – 686. https://doi.org/10.1016/S2212-5671(15)00123-9

50. Winberg, T. M., & Berg, C. A. R. (2007). Students' cognitive focus during a chemistry laboratory exercise: Effects of a computer- simulated prelab. Journal of Research in Science Teaching, 44, 1108–1133.