



OPEN ACCESS

EDITED BY

Regis Kopper,
University of North Carolina at
Greensboro, United States

REVIEWED BY

Marcos Serrano,
Université Toulouse III Paul Sabatier,
France
Bertrand Schneider,
Harvard University, United States

*CORRESPONDENCE

Mina C. Johnson-Glenberg,
✉ mina.johnson@asu.edu

RECEIVED 18 September 2022

ACCEPTED 22 June 2023

PUBLISHED 14 July 2023

CITATION

Johnson-Glenberg MC, Yu CSP, Liu F, Amador C, Bao Y, Yu S and LiKamWa R (2023). Embodied mixed reality with passive haptics in STEM education: randomized control study with chemistry titration.

Front. Virtual Real. 4:1047833.

doi: 10.3389/frvir.2023.1047833

COPYRIGHT

© 2023 Johnson-Glenberg, Yu, Liu, Amador, Bao, Yu and LiKamWa. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Embodied mixed reality with passive haptics in STEM education: randomized control study with chemistry titration

Mina C. Johnson-Glenberg^{1*}, Christine S. P. Yu¹, Frank Liu², Charles Amador², Yueming Bao², Shufan Yu³ and Robert LiKamWa²

¹Department of Psychology, Arizona State University, Tempe, AZ, United States, ²School of Arts, Media and Engineering, Arizona State University, Tempe, AZ, United States, ³Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan, China

Researchers, educators, and multimedia designers need to better understand how mixing physical tangible objects with virtual experiences affects learning and science identity. In this novel study, a 3D-printed tangible that is an accurate facsimile of the sort of expensive glassware that chemists use in real laboratories is tethered to a laptop with a digitized lesson. Interactive educational content is increasingly being placed online, it is important to understand the educational boundary conditions associated with passive haptics and 3D-printed manipulables. Cost-effective printed objects would be particularly welcome in rural and low Socio-Economic (SES) classrooms. A Mixed Reality (MR) experience was created that used a physical 3D-printed haptic burette to control a computer-based chemistry titration experiment. This randomized control trial study with 136 college students had two conditions: 1) low-embodied control (using keyboard arrows), and 2) high-embodied experimental (physically turning a valve/stopcock on the 3D-printed burette). Although both groups displayed similar significant gains on the declarative knowledge test, deeper analyses revealed nuanced Aptitude by Treatment Interactions (ATIs). These interactions favored the *high-embodied experimental group that used the MR device* for both titration-specific posttest knowledge questions and for science efficacy and science identity. Those students with higher prior science knowledge displayed higher titration knowledge scores after using the experimental 3D-printed haptic device. A multi-modal linguistic and gesture analysis revealed that during recall the experimental participants used the stopcock-turning gesture significantly more often, and their recalls created a significantly different Epistemic Network Analysis (ENA). ENA is a type of 2D projection of the recall data, stronger connections were seen in the high embodied group mainly centering on the key hand-turning gesture. Instructors and designers should consider the multi-modal and multi-dimensional nature of the user interface, and how the addition of another sensory-based learning signal (haptics) might differentially affect lower prior knowledge students. One hypothesis is that haptically manipulating novel devices during learning may create more cognitive load. For low prior knowledge students, it may be advantageous for them to begin learning content on a more ubiquitous interface (e.g., keyboard) before moving them to more novel, multi-modal MR devices/interfaces.

KEYWORDS

mixed reality (MR), chemistry, passive haptics, embodied learning, stem education, 3D-printable, XR

1 Introduction

The intersection of educational multimedia and 3D-printed learning objects holds much promise for Science, Technology, Engineering, and Math (STEM) instruction. There are many research-supported design guidelines for 2D multimedia (i.e., using tablets or computer screens) (Mayer, 2009; Clark and Mayer, 2016), and some education-oriented guidelines have been proposed for mobile augmented reality (Dunleavy, 2014; Lages and Bowman, 2019) and for collaborative Augmented and Mixed Realities (AR and MR) (Radu and Schneider, 2019; Radu et al., 2023). However, less research has been done on how learners interact with 3D-printed materials that are tethered to interactive actions on a computer screen (i.e., digitized displays). Designers and creators of educational multimedia content need evidence-based design guidelines that balance the benefits of haptically-mediated platforms with the potential for distraction or cognitive overload.

One of the goals of this study is to understand how users' actions and gestures that are afforded by a manipulable 3D object (in this case, a tangible valve with passive force feedback) affect downstream learning and recall of educational content. Little is known about how gestures with passive haptics affect verbal recall of chemistry content. Passive haptics refers to instances where real objects (also called tangibles) resemble their virtual counterparts and provide feedback through the sense of touch without forces counteracting (Han et al., 2018). This current study focused on the domain of chemistry education, specifically titration. Learners were randomly assigned to either a control or experimental condition. In the control condition they pressed keyboard buttons to release digitized liquid from a digitized burette; in the experimental condition participants could physically turn a stopcock valve on a 3D-printed haptic burette to release the digitized liquid. The burette valve is the passive haptic device in this instance. It necessitates a turning movement which is gesturally congruent to what students do in real world, in-person chemistry labs, thus using this mixed reality device should serve to prime and reify real world skills. Understanding learning outcomes associated with more embodied and gesturally-congruent actions is part of creating evidence-based design guidelines for educational content that is situated along the emerging XR (eXtended Reality) spectrum. The next section describes how media range from Augmented to Mixed to Virtual Reality—XR.

1.1 Focus on mixed reality and haptics

Milgram and Kishino (1994)'s influential paper stated in the 1990s that the Mixed Reality (MR) research field was plagued with "inexact terminologies and unclear conceptual boundaries." To rectify this, they proposed a "Virtual Continuum." At the far left of the spectrum was the real environment (real world), and at the far right was a fully virtual and immersive (digitized) environment, called IVR. In the areas in between where augmentation and overlays could occur, they called that Mixed Reality. The problem of unclear boundaries still plagues researchers today (e.g., How does one classify a VR HMD module that is 60% "pass through" showing the real world?), nonetheless, using the broad term MR allows us to signal that tangible manipulable objects may be used to highlight certain

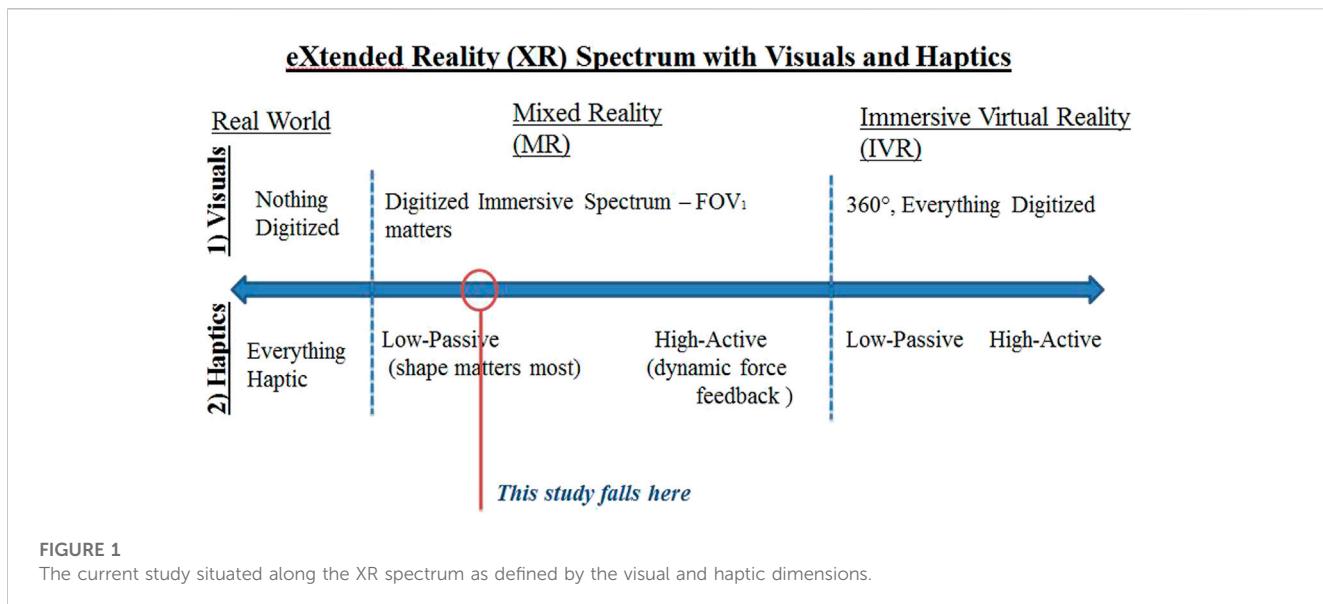
interactive actions in conjunction with digitized components. Our study includes a haptically manipulable object and so it represents class number 6 (p. 1322) of the six classes that Milgram and Kishino (1994) proposed under the umbrella term of MR. Such environments with manipulable objects are also called tangible user interfaces (TUIs); these have matching forms and sizes with objects in the real world and can increase "the digital experience in MR... by helping users perceive that they are physically feeling the virtual content" (Bozgeyikli and Bozgeyikli, 2021).

The haptics in this study are a class of "passive haptics," these devices are solid physical objects that mimic the real world objects and control interaction on the screen. They convey "a constant force or tactile sensation based on the geometry and texture of the particular object" p. 169 (Laviola et al., 2017). Research has shown that passive haptics improve the engagement with the virtual environment (Insko, 2001). A wide range of haptic devices have been developed for virtual reality experiences in general (Bouzbib et al., 2021). Passive haptics in particular have been used in virtual educational contexts to, e.g., teach geography (Yang and Weng, 2016), music (Barmpoutis et al., 2020a), exercise (Barmpoutis et al., 2020b; Born et al., 2020), and kinematics (Martinez et al., 2016). Students using passive haptics in virtual educational contexts are engaging in embodied learning, a type of learning where they use their bodies to touch, feel, pick up, and engage with the devices, which adds another signal to the act of learning (Macedonia, 2019). Haptics can be passive or active. Thus, within the large spectrum of XR there is the shorter spectrum of MR and within the MR spectrum there is a stretch called embedded haptics. Figure 1 is our attempt to parse MR into two dimensions to help clarify where this study sits. The visual dimension is above the arrow, it is primarily based on Field of View (FOV). The haptic dimension is below the arrow and ranges from passive to active. This MR experiment study sits towards the left with low immersivity and passive haptics.

1.2 Embodiment, gesture, and active learning

One of our goals was to design a lesson with a certain amount of embodiment that would use bodily actions or gestures key to learning a specific topic. Proponents of embodiment hold that cognition is inextricably intertwined with our bodies (Wilson, 2002). Bodies interact via environmental affordances (Gibson, 1979) with the world, and humans learn via these interactions (Barsalou, 2008; Glenberg, 2008). The concept of embodied learning is gaining traction in education (Barsalou, 2008; Glenberg, 2010; Lindgren and Johnson-Glenberg, 2013; Johnson-Glenberg et al., 2014).

How can more embodied lessons help students learn? One hypothesis is that learners who are engaged in higher levels of embodiment will learn content faster and in a deeper manner because actions and gestures activate sensorimotor codes that enhance the learning signal. They add a third signal to the usual visual and verbal signals associated with education during encoding and eventually strengthen memory traces. Gestures are a special class of action, and they represent knowledge that is not spoken and they may attenuate cognitive load (Goldin-Meadow, 2011; Goldin-Meadow, 2014). Gestures might do this by "offloading cognition" (Goldin-Meadow et al., 2001; Goldin-Meadow, 2011). Research from Goldin-Meadow's lab has found that requiring children to



gesture while learning a new math concept helped them retain the knowledge; while requiring children to speak, but not gesture, had no effect on increasing learning. Gesturing can play a causal role in learning (Broaders et al., 2007; Cook et al., 2008). The Gesture as Simulated Action (GSA) framework (Hostetter and Alibali, 2019) proposes that motoric pre-planning is another factor that distinguishes gesture (beyond it being a third somatic type of signal along with the typical auditory and visual inputs associated with learning). Per Hostetter and Alibali (2019), “gestures arise from embodied simulations of the motor and perceptual states that occur during speaking and thinking . . . gestures reflect the motor activity that occurs automatically when people think about and speak about mental simulations of motor actions and perceptual states . . . this has also been called co-thought (p. 723).”

We note that for the gesture to aid learning, it should be meaningful and congruent to the content to be learned (Segal et al., 2014; Johnson-Glenberg, 2018; Fuhrman et al., 2020). Johnson-Glenberg (2018), Johnson-Glenberg (2019) proposed three axes in a Taxonomy for Embodied Education; the axes range from low to high and induce embodiment. In such a three-dimensional space the axes can be visualized as 1) magnitude of the gesture, 2) amount of immersivity, and 3) finally “congruence with the content to be learned.” (We note that the turning of the hand gesture to control a stopcock is highly congruent to what chemists do in real labs today.)

The “active learning” literature is not new [see Maria Montessori corpus, and more recent Drigas and Gkeka (2016) for how relatable Montessori’s attention to the hand is with modern digitized content]. Over the decades, many terms have been used to describe gestures and actions from “enactive” to “physical” to “hands-on” to “kinesthetic” to “embodiment.” This article prefers the term embodiment since it connotes a theory as well. Other researchers (see Varela et al., 1991; Nathan and Walkington, 2017; Walkington et al., 2023) cite enactivism as the wellspring. Interesting research on self-performed or enactive tasks demonstrated that when participants were assigned to a self-performing group (“put on the hat”), they recalled more phrases than those who merely heard the phrases (Engelkamp and Zimmer, 1985;

Engelkamp, 2001). Other researchers prefer the term “physical learning.” Kontra et al. (2015) found that students who *physically* held a spinning bicycle wheel on an axle understood concepts related to the angle of momentum significantly better than those who *observed* the spinning wheel. Her participants were then scanned with fMRI while taking the posttest and significant differences were seen in neural activation between the physical and observational groups in the brain’s sensorimotor region. Thus, the amount of “physicality” or embodiment during the learning phase seems to affect brain activity during the assessment phase. There is something about physical haptic input that, *when designed correctly into a lesson*, appears to alter and increase learners’ comprehension. One of the goals of this study is to better understand the boundary conditions of certain educational haptic inputs, and how these haptics might affect different types of learners differently.

In our experiment, we hypothesized that students who interacted with a 3D object that prompted concept-congruent and complex representational gestures during the learning phase of this titration study might learn more. Additionally, we predicted that those who performed the gesture during encoding would use the complex embodied gesture more often during a post-intervention recall. The control condition action consisted of tapping left/right keyboard arrows, although there is some physical activity involved in that action, it was not considered a representational, concept-congruent, or highly embodied gesture.

1.3 Chemistry titration

Definition: In a classic titration experiment, the student learns to use a solution of known concentration to determine the concentration of an unknown solution. Many canonical titration experiments use acid-base neutralization reactions. These acid-base titration experiments are commonly taught in the end of the Chemistry I in high school curriculum (Sheppard, 2006). Students carefully release and mix the known solution from a glass burette (a long cylinder with a valve near the bottom) into a glass beaker containing an unknown

solution until the unknown solution is neutralized. A color indicator that is present in the beaker will permanently change color once the neutralization has occurred. Students can then compute the moles in the unknown solution. They must be able to read a meniscus line to do this precisely. It is critical that the student controlling the drops from the burette with the valve (stopcock) be precise and not “overshoot” for computational accuracy.

1.3.1 XR in chemistry

As VR and AR technologies evolve, their potential to serve the multiple needs of chemistry education continue to be investigated in schools and universities. A variety of XR chemistry laboratory experiences have been developed to teach laboratory experiments and safety, and the field is rapidly expanding (Huwer and Seibert, 2018; Kong et al., 2022). New modules explore the inner workings of laboratory instruments (Naese et al., 2019), and offer a virtual analog of a real experiment (Tee et al., 2018; Chan et al., 2021), all of which can support distance learners. One such VR analog was found to produce no significant difference in learning when compared to a group of students who performed the same experiment in a traditional laboratory (Dunnagan et al., 2020). That study reports a non-inferiority effect, which supports the use of VR as a type of equivalence to an in-class experience. Given that digitized VR is less expensive than in-class learning, many would consider this a positive outcome. Another common application of XR chemistry learning has been the creation of tools that enable simple and complex molecules to be visualized and manipulated in three-dimensional space (Jiménez, 2019; Argüello and Dempski, 2020; Ulrich et al., 2021).

1.3.2 Two examples of previous embodied MR chemistry studies

In this section two previous embodied studies on chemistry that fall along the MR spectrum are described; these are also of interest to highlight the traditional type of analyses that most studies use. This team wanted to use more innovative analyses of comprehension and we present a newer type of Epistemic Network Analysis in the results section. The tests used in the studies below with high school students contained only classic declarative questions and were analyzed with traditional frequentist statistics. This first study contains a condition that falls solidly in the middle of the MR spectrum (see Figure 1). The Wolski and Jagodziński (2019) study used a *Kinect* sensor to map students' hand positions. Their between-subjects study had three conditions: 1) view a film of chemistry lab content, 2) view a teacher's presentation, or 3) actively learn via gesture (gathered with the *Kinect*). In the enactive and embodied *Kinect* condition students were able to manipulate the content on a large projected screen with their hands doing appropriate gestures (note: without haptics and force feedback). Examples include grabbing a virtual beaker or pouring liquid solutions into virtual glassware. On the chemistry posttest, the embodied *Kinect* group consistently out-performed the two other groups (Wolski and Jagodziński, 2019).

Heading to the right of the spectrum, towards the virtual end (with more immersivity and whole-body embodiment), we present a previous co-located classroom experience with very large floor projections. *SMALLab* (Situated Multimedia Arts Learning Lab) is a one-wall CAVE. It includes a floor projection that is 15 × 15 feet. Rigid-body trackable tangible wands were used by the students to

manipulate the virtualized content projected on the ground. Students used the tangibles to pull various virtual molecules into a central, digitized flask while watching the pH level change (all of this projected on the floor). Years ago, a titration study was run in *SMALLab* using multiple high school classes (Johnson-Glenberg et al., 2014). The knowledge assessment format was paper and pencil tests with declarative titration questions. Analyses revealed that those who were randomly assigned to the MR embodied *SMALLab* condition consistently showed more significant learning gains (average effect size of 1.23), compared to those who received regular chemistry instruction (average effect size of 0.24).

1.4 New assessments: Gestures and idea units

One goal of this study was to update the types of assessments that researchers of XR and embodiment are using. The knowledge changes associated with digitized chemistry lessons have traditionally been gathered with written multiple choice questions. This current study moves beyond the declarative and traditional knowledge assessment paradigm and includes a videotaped recall which allows a deeper dive into comprehension via multimodal verbal units and gestures. Gestures and chemistry education have been studied to some degree. At the end of Flood et al. (2015)'s microethnographic investigation, they recommend that students and instructors of chemistry should take more advantage of gestures to create meaning and to interactively negotiate the co-construction of mutual understandings. Flood et al. state that gestures are an especially productive representation in this field because chemistry has the following three attributes, it is: 1) submicroscopic, 2) 3D spatial, and 3) dynamic. Words and drawings do not capture these attributes well. Gestures are “embodied ways of knowing that are performative and ephemeral and are commonly undervalued in favor of texts and written diagrams.” (p.12). Flood et al. hypothesize that there has been little research on the gestures learners make during chemistry education because it is time-consuming to code. We maintain that gestures are a rich and worthwhile source of information and engender insight into a learner's current state of knowledge and are worth the extra effort. There has also been some research on capturing gestures during virtual chemistry labs, see Aldosari and Marocco, 2015.

Gestures: In this study, the recorded recalls were transcribed after the intervention and we tallied both verbal idea units and number of iconic/representational gestures (McNeill, 1992; McNeill, 2008; McNeil et al., 2009). Hostetter and Alibali (2019) state that, while speaking, the likelihood of a gesture at a particular moment . . . “depends on three factors: the producer's mental simulation of an action or perceptual state, the activation of the motor system for speech production, and the height of the producer's current gesture threshold. When there is an active, prepotent motor plan from the simulation of an action or perceptual state, this motor plan may be enacted by the hands and become a representational gesture (p. 723).” The idea of a participant's individual *gesture threshold* is important and it is why statistics in this study were gathered on three other gestures during the recall (besides the stopcock hand-

turning gesture) to demonstrate that the thresholds between conditions were similar.

A recent paper by [Walkington et al. \(2023\)](#) proposes the term MAET for Multimodal Analyses with Embodied Technology. MAET first involves a close examination of learners' videoed gestures, which they define as "spontaneous or planned movements of the hands or arms that often accompany speech and that sometimes convey spatial or relational information." In the classic gesture analysis literature, four classes of gesture are often assessed. There is a class of gestures called rhythmic or "beat gestures"; in our study these were not counted, as they have been shown to have little impact on learning ([Bege et al., 2020](#)). A second class, deictic or pointing gestures, was also deemed out of scope (e.g., it was not semantically meaningful or additive if the participant pointed back to the burette during verbal recall). Thus we focused on two classes of gestures: the iconic and the representational gesture. Iconic gestures illustrate content that is present, or was once, physically present; these gestures often demonstrate a close semantic connection with the speech act. A representational gesture is one that "represents" what the speaker is thinking of, that referent may have *never* been present in the room (e.g., pinching the fingers together to signify droplets of liquid, this is an interesting one that we tallied since the participants never pinched in the experiment, nor did they see a pipette image). Our hypothesis was that the participants in the more embodied condition with the tangible burette would use different gestures during recall compared to the keyboard control group.

The stopcock hand-turning gesture: If gestures are considered "co-thought" ([Hostetter and Alibali, 2019](#)) then they represent a type of knowledge that needs to be scored and honored. In the original, old-fashioned glass stopcocks that chemists used starting in the 1800s, they could literally see with their eyes how their hand-turning very precisely controlled the speed of the drops being released from the burette and into the beaker. Thus, the physical stopcock hand-turning gesture is both an artefact of the original hardware constraints *and* one that is still used in real world classrooms today. In this manner, the stopcock hand-turning gesture is both an iconic and a representational gesture (and perhaps metaphorical). Syncing that classic gesture to a modern digitized device is designed to help prime and train learners for when they get into real labs with real physical burettes and stopcocks. In the end, it is not critical whether the gesture is considered representational, iconic, or metaphorical; we assessed the four key gestures that showed up most frequently throughout all recalls. We demonstrate that only one of them was performed with a significantly different frequency.

Idea units and epistemic network analyses: Idea units are similar to propositions ([Kintsch, 1974](#)). However, for our purposes they do not need to have a "truth value." We scored an idea unit correct if key words were present. Idea units may have included only one key word and that could be either a noun (e.g., burette) or a verb (e.g., reacted).

Because the study has both physical gestures and verbal idea units, it now includes multimodal data. A decision was made to use both Chi-square and Epistemic Network Analysis (ENA) ([Shaffer et al., 2009](#)) to explore and better understand conditional differences with the multimodal data. ENA is based on Social Network Analysis ([Wasserman and Faust, 1994](#)) and it allows for both the idea units and gestures to be in the same model (as opposed to Lag Sequential

Analysis which does not handle co-occurrences). Shaffer et al.'s original ENA was based on the epistemic frame hypothesis which came from a STEM game-playing community. Students played an educational game multiple times over the summer and "cultures of practice" emerged. For ENA, researchers needed to bin or compartmentalize data into nodes. Thus, these practices became the network's nodes. When the nodes were plotted over time, it was evident that an individual's nodes would move over time and the links between nodes would stretch and alter in strength. [Shaffer et al., 2009](#) nodes were practices like knowledge, skill (the things that people within the community do), identity (the way that players see themselves in the community), and epistemology (the warrants that justify players' actions or claims). In our study, our nodes are the individual's idea units and gestures performed as they give a one time recall on titration.

More recently, ENA has been described as a "qualitative ethnographic technique" to identify and quantify connections among elements in coded data and to represent the data in a dynamic network model ([Shaffer et al., 2016](#); [Shaffer and Ruis, 2017](#)). In the recent ENA models, the nodes do not change location over time, but the strengths and weights do change with each data pass. In education, ENA has been used to model and compare various networks, such as regulatory patterns in online collaborative learning ([Zhang et al., 2021](#)), structures of conceptions ([Chang and Tsai, 2023](#)), and metacognitive patterns in collaborative learning ([Wu et al., 2020](#)).

1.5 Research questions

This work is timely and uses novel forms of assessment and statistical analyses to understand learning with MR. We have tethered a 3D-printed tangible that is an accurate facsimile of the type of expensive glassware that chemists would interact with in a real laboratory. One of our goals with embodied MR is to potentially justify the creation of more affordable 3D-printed STEM manipulables for chemistry classrooms. Such printed objects would be particularly welcome in rural and low SES classrooms because glassware is expensive to purchase and effortful to maintain. To understand how learners interact with 3D-printed objects and how those interactions affect learning and retention, **the field needs more randomized control trial (RCT) studies that target variables of interest along the MR spectrum and can unearth causality in learning.** It is important to explore whether, and how, certain gestures performed with haptic devices affect learning and recall, and how prior knowledge might interact with (i.e., moderate) both gesture use and learning. Thus, our five primary research questions are:

RQ1. Knowledge Gains

How does being in the more embodied condition with concept-congruent gestures and passive haptics affect knowledge gains on a traditional chemistry content knowledge assessment?

RQ2. Science Identity/Efficacy

How does one's Science Identity and Science Efficacy predict learning? Additionally, are there any significant aptitude by treatment interactions (ATI), such that level of chemistry prior

knowledge interacts with condition to predict Science Identity/Efficacy?

RQ3. Idea Units recall

How does being in the more embodied condition affect verbal recall of titration? Using both frequentist and ENA Analysis, do we see verbal idea unit recall affected differentially by being in the control versus experimental condition?

RQ4. Gesture

How does being in the more embodied condition with concept-congruent gestures and passive haptics during the learning phase affect use of gesture during testing phase, and also during a recall of titration?

RQ5. Device Interface Preference

At the end of the study, participants were exposed to both interfaces. How does being in a more embodied condition with passive haptics affect preference for type of burette interface: keyboard compared to 3D-printed burette with turnable stopcock?

2 Materials and methods

2.1 Ethics and permissions

The experimental program, all data, and analyses code will be accessible in a public repository 12 months after the final publication of this research. The Arizona State University Institutional Review Board (IRB) approved all experimental protocols under Federal Regulations 45CFR46. All participants signed a consent form. The participants provided their written informed consent to participate in this study.

2.2 Participants

A total of 138 undergraduate students participated for Introduction to Psychology credit (ages 18–29 years old, $M = 19.38$, $SD = 1.88$). However, data from two participants had to be excluded because they received a pretest score of 0; this did not pass our minimum chemistry background check. Thus, the analyses begin with 136 participants. On the study recruitment form, participants saw they should have completed at least two semesters of chemistry in either high school or college. Even though 15 (11.03%) participants reported having taken only one semester of Chemistry, their pretest scores showed adequate Chemistry background. Sixty-eight percent said they had learned about titration. See Table 1.

2.3 Design

The experiment adopts a mixed 2×2 factorial design. The first factor “Embodied Haptic Feedback” is a between-subjects manipulation. Participants were randomly assigned to either the

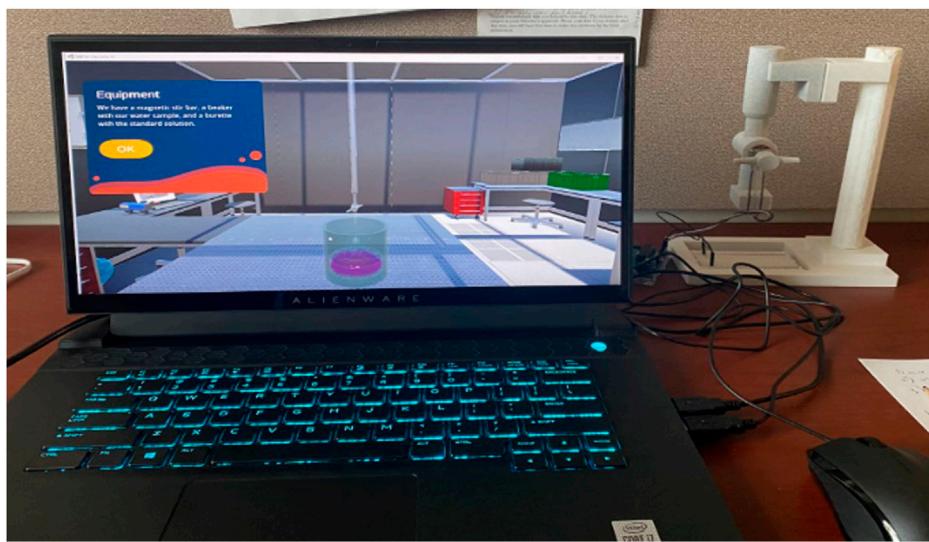
TABLE 1 Participant demographic information.

Demographic information	Number (percentage)
Gender	
Female	52 (38.24%)
Male	79 (58.09%)
Other	4 (2.94%)
Prefer not to say	1 (0.74%)
Dominant hand	
Right	119 (87.50%)
Left	16 (11.76%)
Neither	1 (0.74%)
Ethnicity	
NonHispanic-Caucasian	60 (44.12%)
Asian	36 (26.47%)
Hispanic	27 (19.85%)
Prefer not to say	7 (5.15%)
African American	6 (4.41%)
Semesters of chemistry	
One	15 (11.03%)
Two	60 (44.12%)
Three	16 (11.76%)
Four	29 (21.32%)
More than five	16 (11.76%)
Last observed titration	
Within this month	7 (5.15%)
Within this year	18 (13.24%)
Within last 3 years	50 (36.76%)
More than 3 years ago	17 (12.5%)
Never	44 (32.35%)
Last performed titration	
Within this month	6 (4.41%)
Within this year	12 (8.82%)
Within last 3 years	51 (37.5%)
More than 3 years ago	20 (14.71%)
Never	47 (34.56%)
Grand total	136

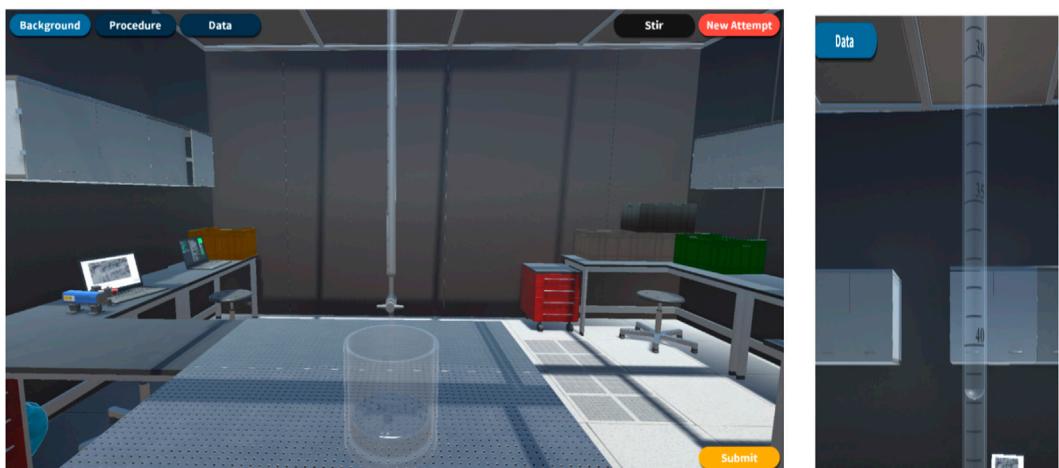
control condition (using a keyboard-controlled virtual burette) or the experimental condition (using a physical 3D-printed haptic burette). The tangible condition is also called haptic-controlled because the fingertips receive tactile feedback when the valve is turned. The second factor is within and uses a repeated measures design. “Test Phase” has two levels: pretest (before the learning phase) and posttest (immediately after the learning phase). Figure 2 shows the computer and 3D-printed haptic burette set-up. Note that the burette is always close to but a few inches from the computer screen.

2.4 Learning materials

Titration simulation module: At the start of every module, the screen displays a short narrative that sets the scene.

**FIGURE 2**

The computer and 3D-printed burette set up. Created with Unity Pro®. Unity is a trademark or registered trademark of Unity Technologies.

**FIGURE 3**

The titration simulation module presented during the learning phase. The left panel shows the virtual stopcock, virtual beaker, and access to the data chart. The tabs at the top show background and procedure information; these are not present in the posttest phase. The right panel shows the meniscus. Created with Unity Pro®.

"You have just begun your new job as a member of a water quality control team for a local government agency. In addition to ensuring that drinking water is clean, you are also responsible for monitoring the acidity of residential lakes. Lakes that are too acidic can be harmful for wildlife." (See [Supplementary Appendix SA](#) under *learning module for full text*).

Participants then proceed to the titration simulation module pictured in [Figure 3](#) where they conduct two repeated trials of one titration simulation task (i.e., the solution in the virtual beaker is identical across the two trials). Participants record data from each of the two trials on a data chart saved to the computer, the

data are used for a posttest calculation question. The center of the module displays a virtual burette with a stopcock at its bottom end and a virtual beaker. To successfully complete a trial of the titration simulation task, participants need to balance the solution in the beaker by adding the solution in the burette (i.e., using the least amount of burette solution possible to turn the beaker solution pink permanently) and then record the amount of burette solution used. Participants either use 1) the left and right keyboard arrows to turn the virtual stopcock to the left or right (control condition) or 2) they kinesthetically turn the 3D-printed stopcock to release the solution from the burette (experimental condition).

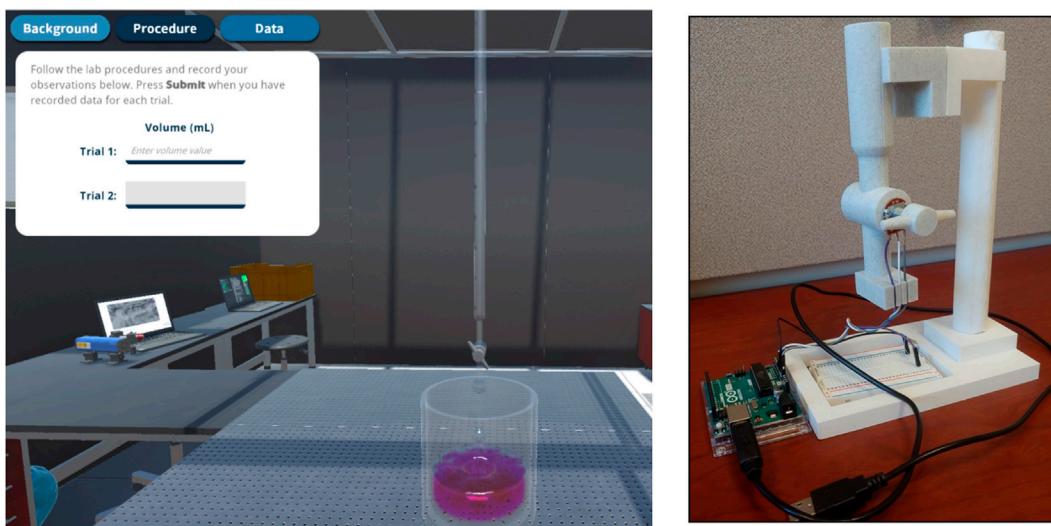


FIGURE 4

Participant view while interacting with the titration simulation module and the 3D-printed haptic burette. The left panel shows what participants see when they turn the stopcock. The further clockwise they turn the stopcock, the faster the solution drops from the burette. The right panel represents a close-up of the 3D-printed haptic burette. Created with Unity Pro®.

The meniscus view (see right panel in Figure 3) the current amount of solution in the burette (participants click “m” on the keyboard to zoom into that). When the solution turns pink at least temporarily they can then click the “Stir” button to activate the mixer bar in the bottom of the beaker. If participants wish to run a new trial and try again, they can click “New Attempt” to restart. To record the amount of burette solution used for each trial of the titration simulation task, participants access the data chart by clicking the “Data” tab. They type in a number and click the “Submit” button. This concludes a trial.

Participants use the titration simulation module three times total. They use it once during the learning phase. They then use it twice more during the posttest to demonstrate their ability to conduct titration experiments. The module is identical from the learning phase to the posttest phase, except for two aspects: the control with which participants use to turn the stopcock of the virtual burette (see subsection “Controlling the Virtual Burette Stopcock” below) and the “Background” and “Procedure” tabs, which are only available during the learning phase. Two subject matter experts (one a high school STEM teacher and one a University chemistry professor created the texts explaining the background of the titration experiment (see Supplementary Appendix SA) and the texts breaking down the procedure of the titration experiment (see Supplementary Appendix SB); they also co-designed the content knowledge tests.

2.5 Apparatus

The experiment was conducted on an *Alienware* laptop (model m15 R4). During the experiment the mouse and 3D-printed haptic burette were positioned (on the left- or right-hand side) of the keyboard according to participant hand dominance. The 3D-printed haptic burette consists of a cylindrical burette section, a stand, a base, a

stopcock lever, a breadboard, a potentiometer, a resistor, three male-to-female wires, a male-to-male wire, an *Arduino Uno*, and a type A USB cable (see Figure 4B for an assembled 3D-printed haptic burette). We used a 3D printer (*Ender-3 3D Printer—Creality*) to print the burette, the stand, the base, and the stopcock with plastic filament, PLA. Turning the stopcock changes the potentiometer resistance. The 3D-printed stopcock is gesturally well-mapped and turning it is congruent to turning a classic glass stopcock used in real world labs.

2.5.1 Experimental manipulation—manipulating the stopcock

In both conditions, the further clockwise that the participants turn the virtual burette stopcock, the faster the virtual solution drops from the burette. There are four total droplet speed options that accelerate exponentially. During the titration simulation module, participants either control the 1) virtual stopcock on the virtual burette by using the *keyboard-controlled burette* (i.e., keyboard left and right arrow key moves the virtual stopcock on screen) or 2) the *3D-printed haptic burette* (i.e., turning the 3D-printed stopcock pictured in Figure 4 which also moves the virtual stopcock on screen). During the learning phase, participants in the control condition use the keyboard-controlled burette (note: the 3D-printed haptic burette is hidden under a cloth from control condition participants’ view during the learning phase), and participants in the experimental condition use the 3D-printed haptic burette.

During the learning phase, only the experimental group interfaced with the 3D-printed burette. However, during the testing posttest phase, there were two further trials of titration simulation and *both* types of burette control mechanisms were used. These occurred in a set order. For the *first* posttest trial, all participants used the keyboard-controlled burette. Thus, this interface was novel to the experimental condition. For the *second* trial, all participants used the haptic-controlled 3D-printed haptic burette, thus, this interface was novel to the control condition. We

needed all participants to try both interfaces so we could ask the within-subjects interface comparison question at the very end of the study. The control condition finally sees the 3D-printed haptic burette during the posttest phase when the covering cloth is taken off of it. The experimenter always instructs the participant on how to control each type of burette device whenever it is first presented. The burette is placed near whichever hand the participant claims is dominant.

2.6 Measurement and assessment items

Descriptions for pre- and posttest items are provided below, but see [Supplementary Appendix SC](#) for the complete set of measurement and assessment items listed in order as they appear in the experiment.

2.6.1 Science identity measures

To explore how an embodied chemistry learning experience may affect participants' self-view as a chemistry/science student and *vice versa*, several self-report items were included at both pretest and posttest phases. [Stets et al. \(2017\)](#) was referenced during development of these items. A total of three items were included on participants' identity as a chemist, and four items on participants' perceived scientific efficacy (See [Supplementary Appendix SC](#)).

2.6.2 User experience measures- posttest only

The effect of embodied learning on participants' learning outcome likely covaries with participants' user experience with the titration module. Thus, we include items during the posttest phase for participants to self-report their user experience. These items cover subjective aspects of enjoyment, engagement, and ease of use. The item on enjoyment was adapted from [Lowry et al. \(2013\)](#), three items on how much the virtual titration application engaged participants were adapted from [Schaufeli et al. \(2006\)](#), and four items asking about the ease of use of the burette as well as participants' preference between the keyboard-controlled burette and the 3D-printed haptic burette were adapted from [Davis \(1989\)](#).

2.6.3 Chemistry knowledge assessment items

Chemistry knowledge items were created by two chemistry instructor SMEs. There are 12 items in the pretest in the formats of multiple choice, drag-and-drop, numerical response, and textual free-response. Of these 12 items, seven were general and five were titration specific. For posttest the assessments were split between seven *repeated* items that were seen on the pretest (five of which were general and two were titration specific) and eight new items (four general ones and four titration specific ones). Among the new titration specific items are three calculation questions based on saved data charts created by the participants during the three titration simulation tasks.

2.6.4 Verbal recall and gestures

A recorded recall was the very first post-intervention task performed. We were also interested in assessing students' comprehension of titration beyond a text-based metric. The prompt spoken by the experimenter and present on the laptop screen was:

"Pretend you are a teacher. How would you describe a titration experiment to a student."

The experimenter ensured the student's chair was moved back at least 5.0 feet (1.52 m) from the laptop camera. The laptop recorded both audio and video. The goal was to make sure that hand gestures and actions were in the camera frame.

Coding and scoring scheme for idea units and gestures: To conduct the video analysis, a coding scheme was vetted by, iterated upon, and finally agreed to by two SME's and two graduate students. All rubric designers were blind to condition. See the idea units table placed in the results section for the 10 units that were chosen. The 10 canonical idea units needed to be present for an "expert-like" recall of how titration is done. (Three expert videos from science teachers were assessed to create an expert recall.) Each unit was worth one point. Additionally, there were four frequent gestures (non-beat and non-deictic) that encoded information beyond the verbal and these were scored as well.

2.7 Procedure

The experiment consists of three phases: pretest, learning phase, and posttest phase. Participants progress at their own pace, and the entire duration of the experiment is approximately 90 min. During the pretest phase, participants provide demographic information, respond to science identity measurements, and then complete the chemistry assessment items. During the learning phase, participants read background text about titration and then use the titration simulation module to complete two trials of the titration simulation task. During the posttest phase, participants complete the video-recorded recall response, self-report their learning outcome, respond to user experience measurements, respond to science identity measurements, provide general feedback, complete 12 chemistry assessment items, and then perform two titration simulation tasks followed by two calculation questions based on their self-created data charts. [Figure 5](#) shows the flow of the experiment.

3 Results

The results are listed in the order of the research questions posed at the end of the Introduction section.

3.1 RQ1: chemistry and titration knowledge

The first research question addresses knowledge gains. These have been split into two categories:

- 1) *General chemistry only* (pretest $n = 7$, posttest $n = 9$).
- 2) *Titration specific only* (pretest $n = 5$, posttest $n = 6$).

3.1.1 Matching at pretest

General chemistry: For the knowledge gain analyses, data from two participants were excluded due to missing data. To ensure that participants' chemistry background knowledge did not differ between the two conditions, independent samples *t*-tests were run on the two types of knowledge: Pretest *general* chemistry

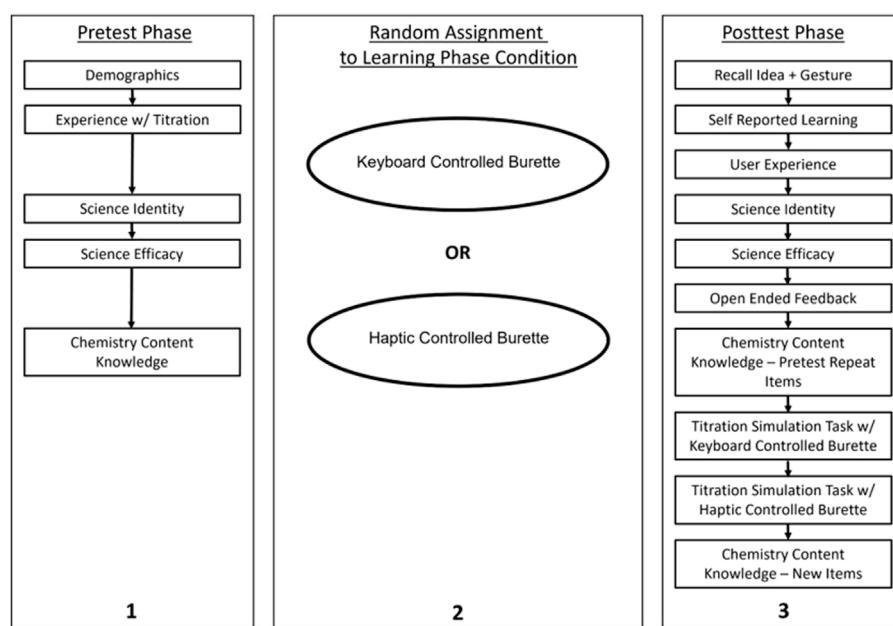


FIGURE 5

Flow of the procedure. Participants are randomly assigned to use either the keyboard-controlled burette (control condition) or the 3D-printed haptic burette (experimental condition).

TABLE 2 Means and standard deviations for five of the key variables in this study.

Key variables	Mean		Standard deviations	
	Control	Experimental	Control	Experimental
Posttest General Chemistry	7.84	7.78	3.15	3.4
Posttest Titration Specific	3.49	3.64*	2.79	3.8
Idea Units	4.78	4.84	1.95	1.92
Gesture Stopcock-turn in Recall avg.	0.98	1.38*	2.01	1.87
Device Preference	1.38*** ^a	1.40*** ^a	1.50	1.52

^aBoth groups significantly preferred using the 3D-printed burette, these results test whether groups differ from a midpoint of 0.

Sig. codes: If p is less than or equal to “*” 0.05, *** 0.001.

knowledge and Pretest *titration* specific knowledge. Condition was the independent variable. Scores on the general chemistry knowledge pretest (maximum score of 15 points) did not significantly differ: Control ($M = 5.88$, $SD = 2.35$), Experimental ($M = 5.63$, $SD = 2.54$), $t(133.13) = 0.60$, $p = 0.55$, 95% CI $[-0.58, 1.08]$.

Pretest titration specific knowledge: These questions (maximum score of 15 points) did not differ between the control condition ($M = 2.18$, $SD = 2.82$) and the experimental Condition ($M = 2.50$, $SD = 3.05$), $t(133.17) = -0.64$, $p = 0.52$, 95% CI $[-1.32, 0.67]$. Thus, the conditions were well-matched.

3.1.2 Posttest knowledge questions

Posttest general chemistry: A multiple linear regression predicted accuracy on posttest general chemistry knowledge questions (maximum score of 17 points) with Condition (coded as “-1” for

control condition and “1” for experimental condition), and a covariate of accuracy on pretest general chemistry knowledge questions (grand mean centered). Pretest predicted post, $B = .90$, $SE = 0.09$ $t = 10.38$, $p < 0.001$. Condition was not significant, $B = 0.06$, $SE = 0.21$, $t = 0.30$, nor was the interaction term of condition by pretest. Table 2 presents the Means and Standard Deviations for five of the key variables in this study.

3.1.3 Titration specific questions

A multiple regression was run with posttest titration specific knowledge questions (maximum score of 17 points) as the DV. The analysis included a covariate of accuracy on pretest titration specific knowledge questions (maximum score achieved of 15 points, grand Mean centered), and their interaction. The overall regression was statistically significant [$R^2 = 0.39$, $F(3, 130) = 28.18$, $p < 0.001$]. Table 3 lists the statistics.

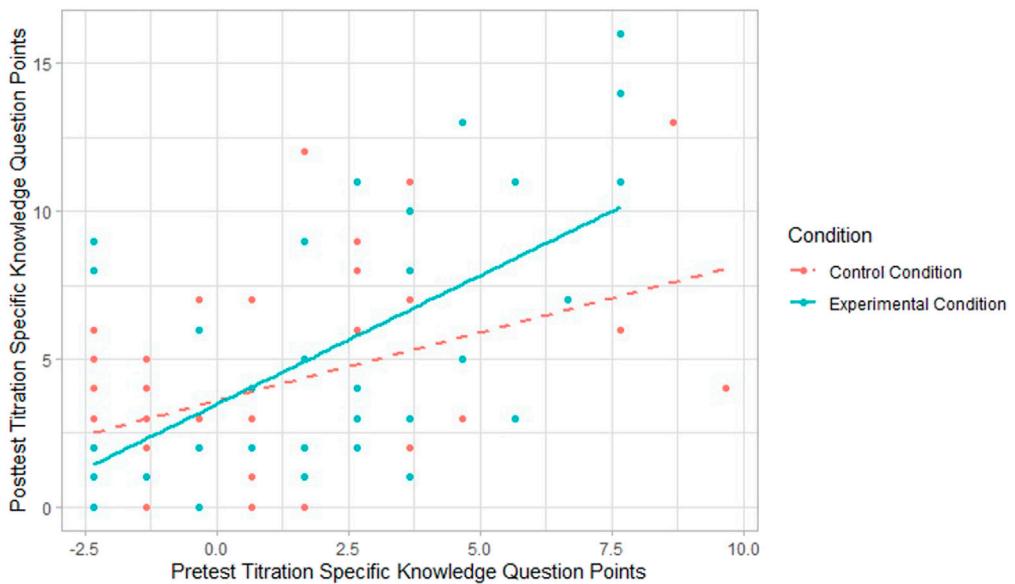


FIGURE 6

Interaction between condition and performance on pretest titration specific knowledge questions, the DV is posttest titration specific knowledge.

TABLE 3 Statistics for multiple regression coefficients predicting posttest titration specific questions.

Coefficients	Estimate	Std. Error	t value
(Intercept)	3.53	0.23	15.56***
Condition	-0.06	0.23	-0.26
Pretest titration	0.67	0.08	8.56***
ConditionXPretest titration	0.20	0.08	2.63**

Sig. codes: If p is less than or equal to “*” 0.05 “**” 0.01 “***” 0.001.

Bolded values are statistically significant.

The intercept is significant when Mean-deviated to 3.53 pretest points. Pretest titration specific questions significantly predicted posttest titration specific knowledge questions. The interaction term of condition by pretest is significant. This aptitude by treatment interaction (ATI) implies the groups are answering these questions differently depending on their prior knowledge AND group placement. Figure 6 shows the interaction; participants in the experimental condition who have average and higher pretest scores answer the titration specific posttest questions better.

To highlight in one graphic the variables of interest and how the groups differed, we have created Figure 7. The chemistry and titration knowledge differences are the first two variables in 7A. Device preference is described at the end of the results section.

3.2 RQ2: science identity and efficacy

The second research question concerns participants' self-reported Science Identity and Science Efficacy. Two analyses here

focus on whether these were 1) predictive of posttest knowledge scores, and 2) did the identity or efficacy scores at post-intervention significantly differ depending on condition.

3.2.1 Pretest match

To ensure that the two conditions did not differ in participants' science identity and science efficacy background, we first conducted two Wilcoxon-Mann-Whitney tests each with dependent variables pretest science identity (maximum of 21 points) and pretest science efficacy (maximum of 28 points) and condition as the independent variable. Results from the Identity test show that the median pretest science identity ratings for the control condition of 7 ($IQR = 4$) did not significantly differ from the median in the experimental condition of 6 ($IQR = 5$), $p = .95$. Similarly, results from the Efficacy test did not show a significant difference between pretest science efficacy ratings for the control condition ($Mdn = 21.5$, $IQR = 5$) and the experimental condition ($Mdn = 21$, $IQR = 4.25$), $p = 0.99$.

3.2.2 Science identity predicting posttest knowledge

This multiple linear regression predicted accuracy in posttest repeated questions with pretest science identity ratings (grand mean centered), Condition (coded as “-1” for control condition and “1” for experimental condition), and their interaction. The DV included all knowledge questions both general chemistry and titration specific. The fitted regression model was statistically significant, $R^2 = 0.09$, $F (3, 130) = 4.30$, $p < 0.01$. See Table 4.

Pretest science identity was not predictive. But, condition did significantly predict accuracy in posttest knowledge questions; additionally, the ATI of pretest by condition was significant. This means that participants in the experimental condition showed better

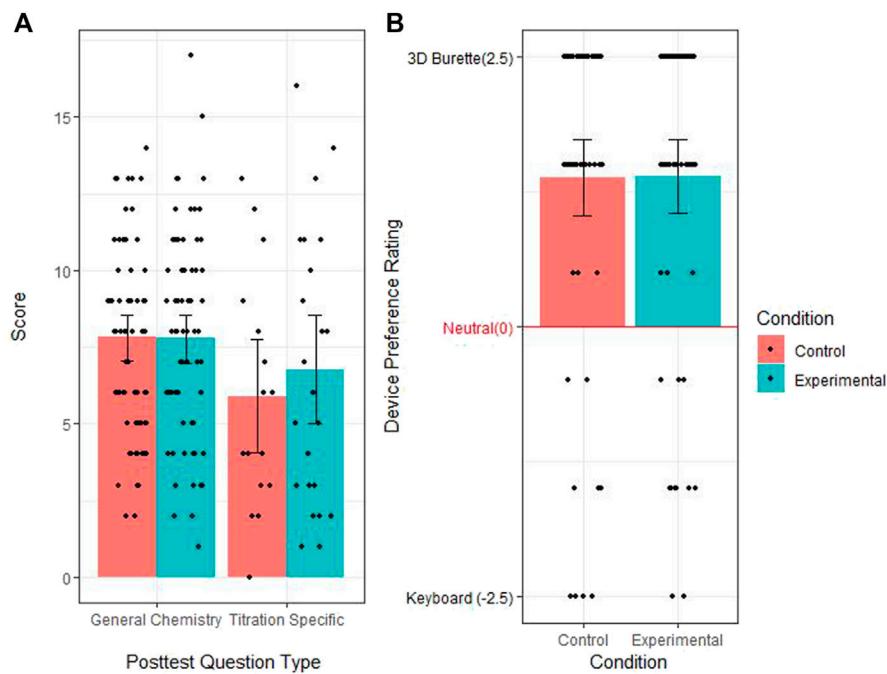


FIGURE 7

Graphs of interest: (A) shows the Performance on Knowledge Questions and (B) shows Device Preference by Condition.

TABLE 4 Statistics for science identity multiple regression coefficients predicting posttest questions.

Coefficients	Estimate	Std. Error	t value
(Intercept)	6.22	0.23	27.54***
Pretest Science Identity	0.03	0.23	0.13
Condition	0.15	0.06	2.52*
Pretest Science IdentityXCondition	0.14	0.06	2.29*

Sig. codes: If p is less than or equal to “*” 0.05 “**” 0.01 “***” 0.001.
Bolded values are statistically significant.

chemistry knowledge retention compared to participants in the control condition *as their science identity ratings increased*. This makes sense because Science Identity and posttest scores are correlated ($r = 0.23$ $p < 0.01$). See Figure 8 for a plot of the interaction.

3.2.3 Science efficacy

The second multiple linear regression predicted accuracy in posttest repeated questions with pretest science efficacy ratings (mean deviated) and condition. The regression was just significant [$R^2 = 0.05$, $F (2, 131) = 3.58$, $p < 0.05$]. The same pattern was seen as with Science Identity. Condition was only a significant predictor when it was embedded in the ATI interaction of conditionxpretest, showing again that those in the experimental group with higher efficacy scores did better on the knowledge test compared to those in the control condition. Note: Science Identity and Efficacy change scores from pre to post-intervention were calculated and those are included in Supplementary Appendix SC, the change scores did not differ by condition.

3.3 RQ3 and RQ4: recall—idea units and gestures

Videos were recorded of all students, but some technical difficulties were encountered, e.g., on seven of the videos the chair was misplaced and hands could not be seen, those videos were not included. The following analyses were run on 54 controls and 58 experimental participants.

Table 5 shows the types of idea units and gestures that were tallied. Table 6 shows the means and counts for the idea units and the gestures. Idea units were normally distributed across groups and a t -test revealed they were not significantly different between groups, $t (110) = 0.18$, $p = 0.86$.

3.3.1 RQ3: regression on idea units

A multiple regression was run with idea units as the DV with pretest knowledge as a predictor, it also included the pretest knowledge by condition interaction. The overall regression model was statistically significant, $F (3, 108) = 6.77$, $p < 0.001$. This is because the constant (intercept) parameter was significant ($B = 7.40$, $p < 0.001$), however, none of the other variables were significant predictors {only pretest knowledge was a trend, [$t (110) = 1.79$, $p = 0.08$]}.

3.3.2 RQ4: chi square on gesture units

How does being in the more embodied condition affect use of gesture during recall of titration? As seen in the bottom of Table 6, four of the most frequent gestures were tallied: G1 Pinch, G2 HandCurl—the burette gesture which referenced the long cylinder, G3 HandTwist—dynamic stopcock turning gesture, and G4 Stir. A Chi-square analysis revealed that only G3, the stopcock hand-turning gesture was performed significantly more often by the experimental group, G3: $\chi^2 (1, N = 112) = .58$, $p = 0.016$.

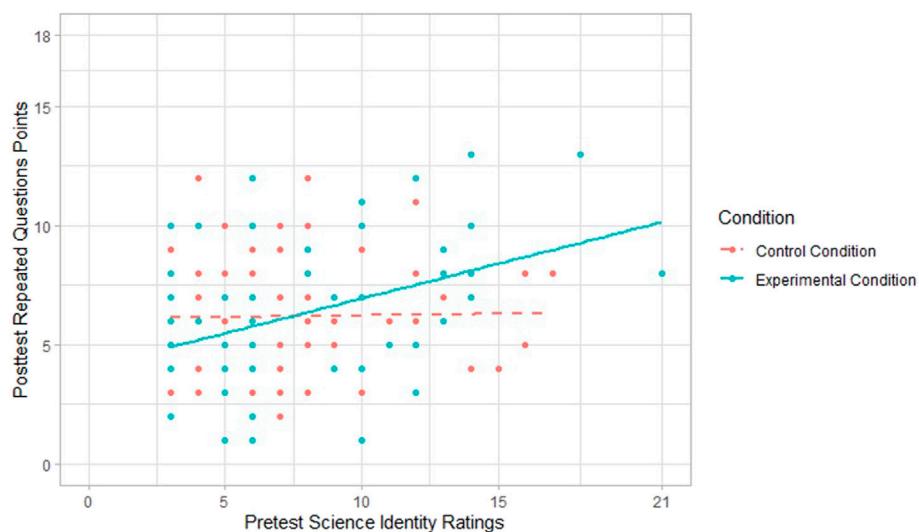


FIGURE 8
Interaction between condition and science identity.

TABLE 5 The descriptions and counts of idea units and gestures.

Idea units	Description	Label	Example
I1	Concentration	Concentration	"In titration you have a concentration of a solution; that is unknown" (or known)
I2	Acid and bases	AcidBase	Mention of either acid or base
I3	The indicator (or phenolphthalein)	Indicator	Mention of indicator (only two specifically said "phenolphthalein")
I4	Color changes of the solution	ColorChange	Color change or "turns pink" when added
I5	React, or reagent during experiment	Reagent	Mention adding enough to have it react
I6	Precision manipulation during experiment	FineManipulate	Add little by little ("drop by, drop, slowly")
I7	Experiment equipment (not stopcock)	Equipment	Generally burette (or "tube, cylinder", etc.)
I8	Stopcock related	Stopcock	Specific to stopcock (or "dial, valve, thing")
I9	Experiment termination conditions	StopChange	When permanently pink, stops changing color, stirred
I10	equation/math language	Math	Math or mention equation, i.e., "volume of base to figure out moles, one to one ratio, divide number of molecules by the volume of the acid for concentration..." etc.
Gesture units	The hand is	Label	
G1	pinching to represent drops	Pinch	
G2	Curling, either with up-and-down motion or not, represents burette	HandCurl	
G3	Twisting to the right or the left, represents stopcock	HandTwist	
G4	Flat, or index finger is dipping downwards and going in a circle, represents stirring, twirling	Stir	

The three other gestures were closely matched between groups: G1: $\chi^2 (1, N = 112) = 0.28, p = .61$; G2: $\chi^2 (1, N = 112) = 0.01, p = 0.94$; and G4: $\chi^2 (1, N = 112) = 0.04, p = .84$. We report these three NS gesture statistics to show that groups were very matched on how much they gestured and used their hands overall during recall. This establishes the baselines or set point

(average frequency) between the two groups. Set points vary by culture (among other variables, [Hostetter and Alibali, 2019](#)), thus, it is important to make sure the groups were matched, so we can make precise claims regarding the difference associated with the one manipulated variable, i.e., the hand-turning stopcock gesture.

TABLE 6 Descriptives for the idea units and counts for the gestural units.

Idea units	Control			Experimental		
	N	Mean	SD	N	Mean	SD
	54	4.78	1.95	58	4.84	1.92
Gestures	Sum			Sum		
G1 (Pinch)		18			22	
G2 (HandCurl)		19			20	
G3 (HandTwist)		15*			29*	
G4 (Stir)		12			12	
Gestures sum		64			83	

* Asterisk signifies that the values are significantly different from each other.

3.4 Epistemic network analysis (ENA) on recall

The ENA web tool (<https://app.epistemicnetwork.org>) was used to model and compare students' recall patterns. An ENA was run using the 10 verbal units and the four most common representational gestures given during the recall. There were technical difficulties recording and accessing 22 videos from the full sample of 136. These losses were equally distributed across both conditions. Additionally, two participants misunderstood the directions. When prompted to describe a titration experiment, they proceeded to describe *the* experiment they were currently in and started with the following, "First, I signed the consent form. Then I sat at the computer..." These two were excluded from the analyses, leaving this section with 112 participants.

In ENA analyses, each idea unit and gesture unit can be conceptualized as a node. The weight or the links between nodes symbolizes how related (co-occurring) the node is in each individual's ENA. Thus, a network of ENA is made for each participant. To do this, ENA first creates an adjacency matrix for the 14 nodes. The algorithm then normalizes the matrix and performs SVD (singular value decomposition) on the data. It then performs a sphere or cosine norm on the data and centers it (without rescaling). The centroid of a network is similar to the center of a mass, or sphere. The arithmetic mean of the network is based on the edge weights as well.

We have created two group-based ENA's. In **Figure 9**, the color red denotes the low embodied condition (control) and blue denotes the high embodied condition (experimental). Each dot represents a participant and where their ENA centroid would land in the two-axis projection of the ENA structure.

Figures 9A–C ENA networks of the recalled idea and gestural units. (a) Control ENA network. (b) Experimental ENA network. (c) Subtracted network of two groups.

The nodes in a projected ENA do not move around so comparisons can also be made visually. Note that I4 (ColorChange) and G2 (HandCurl) are the extremes of the Y axis; note also that I2 (AcidBase) and G3 (HandTwist) are the extremes of the X axis. These set the boundaries of the ENA. In multidimensional projections, the locations themselves are not always interpretable.

What is of interest is *if, and how far, the groups differ when overlayed onto the same projected network*. **Figure 9C** shows the comparison network of those in both groups (by subtracting their networks). There is a discernible difference between the two recalls.

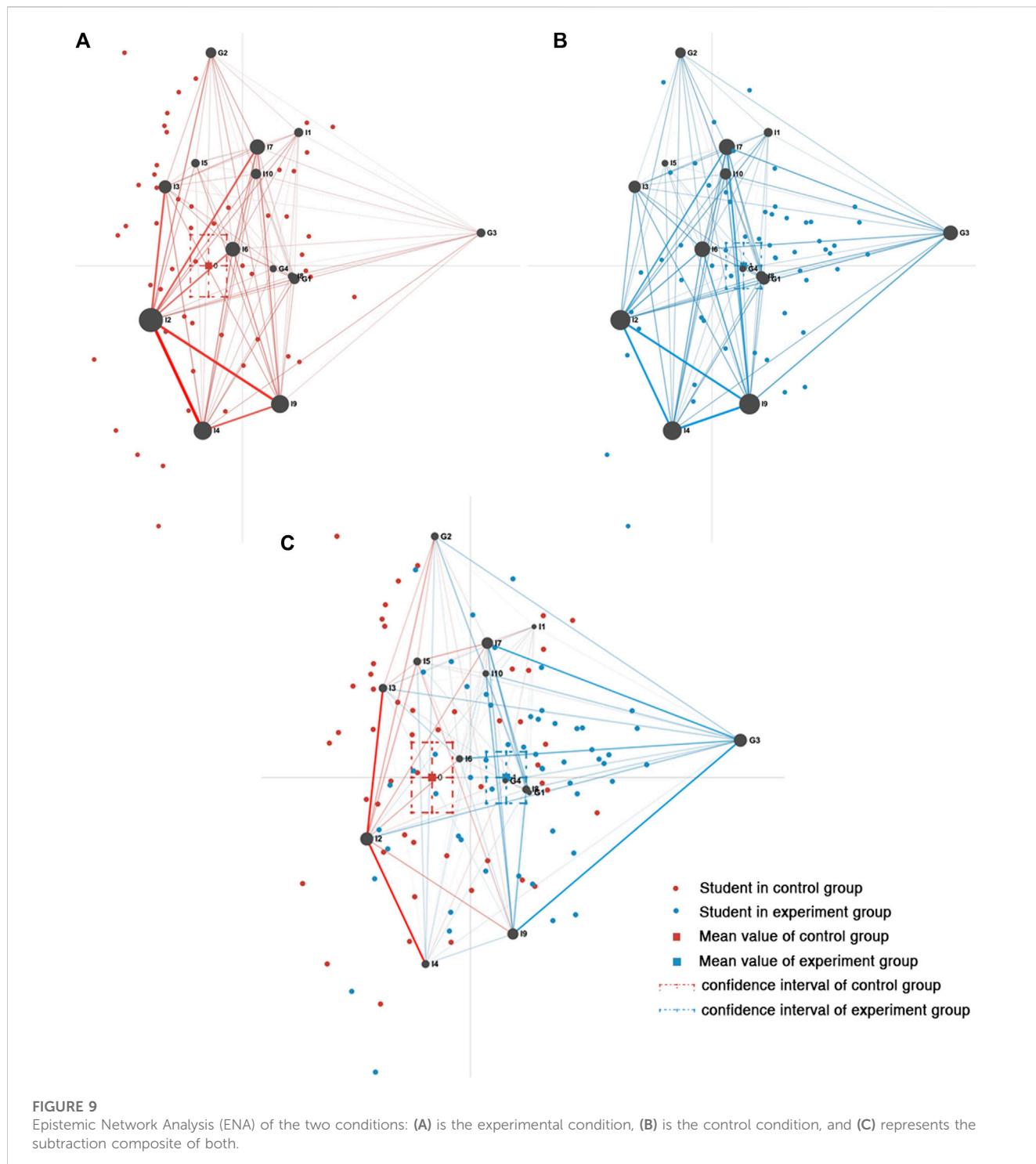
In an ENA graph, the thicker the line, the stronger the connection between two nodes. In **Figures 9A, B**, the two groups both shared strong connections and a triangle is created between I2 (AcidBase) and I4 (ColorChange) and I9 (TermiCod)—these are all verbal outputs. However, when the graphs are subtracted and we can see the differences (**Figure 9C**) certain saliences emerge. Specifically, students in the control group used far fewer gestures. The high embodied experimental group is expressing their comprehension of titration by also using "nonverbal knowledge"; if gesture is considered co-thought, then more knowledge is being reported by the experimental group when recall is coded in a multimodal manner.

Inspecting the strongest links in the subtraction ENA of 9c, we see the highest values in the experimental group represent a triangle of I7 (Equipment) and G3 (HandTwist) and I9 (TerminateCond-stop color change). G3 (HandTwist) played an essential role in connecting the ideas and gestures in the experimental group, but it is very weakly linked for the control group. A frequentist analysis can be done on the distribution of projection points between the low and high embodied groups. An independent *t*-test was performed using each student's ENA on the two axes. A *t*-test on the two centroids shows that on the Y-axis there was no difference between groups, but on the X-axis the centroid of the experimental group was significantly further to the right compared to the control group, [t (109.57) = 5.19, $p < 0.001$, Cohen's $d = 0.98$]. Visual inspection confirms that the epistemic network for the experimental group is significantly affected, or altered, by their use of the hand-turning stopcock gesture (G3).

3.5 RQ5: device interface preference

At posttest, all participants experienced both the 3D-printed haptic burette and the keyboard-controlled burette to control the stopcock. At posttest a survey item included a digital slider that allowed participants to choose which device they preferred. There were six positions along the scale, and the keyboard-controlled burette was located on the left end (next to the first position with the numeral 1); the 3D-printed haptic burette was on the right end [next to the numeral 6 (see **Supplementary Appendix SC** for question format)]. Participants were instructed to position the slider closer to the device they preferred. In the output, a rating of 1 indicates a strong preference for the keyboard, and a rating of 6 indicates a strong preference for the haptic device. For the following analysis, we centered these ratings at 3.5, the center point of the scale which theoretically denotes no preference. After centering, a negative rating indicates a position to the left of the center point and a positive rating indicates a position to the right of the center point (i.e., towards the 3D-printed haptic burette).

A one-sample *t*-test was conducted. Results indicated that overall ALL participants preferred the burette, (control $M = +1.38$, $SD = 1.50$) as well the experimental group ($M = +1.40$, $SD = 1.52$). Both groups were significantly different from neutral (or the 0 point). That is, a significant majority, regardless of condition, placed the slider towards the positive direction from the 0 center point, t (133) = 10.68, $p < 0.001$, 95% CI [1.13, 1.65].



4 Discussion

This study adds to the evidence base regarding how students learn from, and interact with, the new class of mediated, 3D-printed mixed reality devices. We focus on a passive haptic device that allows users to mimic a key action performed with real lab glassware. The experimental manipulation centered on kinesthetically turning a stopcock lever to control a tangible burette, as compared to a keyboard button-press condition, which would feel like

“business-as-usual” for those in online chemistry labs. Online labs are an important and growing segment in education, but the “real” kinematic feeling of manipulating glassware and lab equipment is lost. 3D-printed devices may serve as cost-effective bridges between the real glassware experience and the purely digitized desktop experience. Such cost-effective devices may be especially useful for rural or low SES communities. To that end, this experiment explored five key research questions and each discussion follows below:

4.1 RQ1: science knowledge effects

Of primary interest to this team were conditional effects on posttest science knowledge. The questions were divided into general chemistry knowledge and titration specific questions. There were no significant conditional main effects on the general chemistry knowledge questions. Several of the important statistically significant results in this study were not found with main effects, but were discovered by unearthing nuances in higher-order interactions. A significant interaction on the titration specific knowledge questions was seen and suggests that *improvement from pretest to posttest depended not only on which condition the learners were in, but also on how much prior chemistry knowledge they possessed*. The interesting aptitude by treatment interactions in this study consistently favored the higher prior knowledge students who were in the haptic experimental condition.

We contend that using a novel, MR 3D-printed device was most beneficial for those with average to high prior science knowledge. Our passive haptic, multi-modal experience presented an interface that is not seen in the usual computer set-up that students would have experienced. The 3D-printed device not only looks (and feels) different, it was located further from the display screen than the keyboard. Though this was only inches away, we presume some cognitive effort associated with visual integration needed to occur. Additionally, touching and turning the stopcock provides a separate channel for haptic feedback. It may be the case that *adding another sensorimotor channel of feedback for the lower prior knowledge students increased their cognitive load and that proved deleterious for learning the titration content*. Lower prior knowledge students may already have been struggling with resource allocation issues and integrating the feedback from the haptic device may have been an additional signal that was difficult for them to assimilate. In addition to a resource explanation, it may be that some lower-knowledge students need to have access to previously stored touch sense modalities in order to integrate touch with the abstract concepts. A meta-analysis by [Zacharia \(2015\)](#) comes to that conclusion about the importance of “touch sensory” feedback in STEM education when comparing physical manipulation conditions to virtualized content. Their research suggests that:

“...touch sensory feedback is needed when the knowledge associated with it is incorrect or has not been constructed by the student through earlier physical experiences... The fact that touch sensory feedback appears to be needed when the knowledge associated with it has not been constructed by the student in the past, complies with what embodied cognition is postulating, namely sensorimotor input made available during previous experiences is stored in memory and reused for future internal cognitive processing.”

More research needs to be done on the causation of these sorts of aptitude by treatment interactions. To support the cognitive load conjecture, future experiments should include biometric measures of cognitive effort and load.

In terms of application, one suggestion for instructors may be to start the lower prior knowledge students off on a more traditional and ubiquitous interface (e.g., keyboard) and then when knowledge is more stable and integrated move the students to the more tangible and haptic interfaces that better simulate real chemistry equipment. In this manner, these nuanced ATI interactions can inform the decisions of

educators of online science courses as they weigh options for delivering the laboratory component of courses. If a virtual version of the experiment is chosen, educators must decide if a mouse-and-keyboard option is appropriate, or if it is worthwhile to invest in a haptic interface that enables students to perform the same gestures that students would in real experiments. Given the costs associated with purchasing peripherals and the results of this study, educators may want to wait for a knowledge threshold (mastery) to be met before introducing a 3D-printed MR burette interface to students. For example, students in introductory chemistry courses will presumably have lower prior knowledge and might not perform as well with the haptic burette device. Students in the more advanced chemistry courses might optimally benefit from being required to perform the turning gesture characteristic of titration on a haptic device and this would increase their knowledge gains. One takeaway is that we recommend some scaffolding or “easing into” for lower prior knowledge students before they are exposed to more novel interfaces and MR devices.

4.2 RQ2: science identity and efficacy

Science identity: At pretest, science identity ratings significantly correlated with all knowledge test variables and with pretest and posttest efficacy (*except* for posttest titration specific questions). There were no significant main (conditional) effects on science identity, but we did see more evidence of the nuanced ATIs. Posttest knowledge questions were answered differently depending on level of pretest science identity *and* which condition students were in. The interaction revealed that those with a higher pretest science identity did significantly better on the posttest questions. This suggests that when students identify strongly as scientists *and* use a haptic device that simulates what is in a lab, then those students will do better on the titration questions. Note: Science identity and science efficacy were significantly correlated with each other (pretest $r = 0.42$, test correlations are in [Supplementary Appendix SC](#)).

4.3 RQ3: recall of idea units

4.3.1 Regressions on idea units

Students were prompted to pretend they were teachers and to describe how titration is done in a lab. The range of individual students’ verbal recall idea units was 1–12. Notably, the average number of idea units recalled by both conditions did not differ. Participants talk about the experiment in the same way. In the regression on idea units, only pretest knowledge about chemistry was a statistical trend predicting verbal idea unit recall. Traditionally, assessing knowledge has been done verbally, either with text-based questions or verbal speech recalls. That is certainly one type of “knowledge”; however, we believe that when researching novel haptic interfaces (especially ones associated with unique gestures like a hand turning), it is important to also gather another type of knowledge unit called the *gesture unit*.

4.4 RQ4: gesture units

The most frequent and salient videotaped gestures were coded, tallied, and analyzed. The two conditions performed three of the

four salient gestures very similarly, i.e., pinching gesture to connote slow release of liquid, the “curled palm” burette gesture, and the twirling of the finger gesture to connote stirring. This confirms that the baseline, or set point (Hostetter and Alibali, 2019), of gesture for the conditions was similar. We had predicted that the experimental group, which physically and kinesthetically turned the 3D-printed stopcock during the learning phase, would use the “hand turning” gesture (G3) significantly more often during recall. This was confirmed in our multi-modal analysis and it represents a different kind of knowledge output. Using the 3D-printed device in the learning phase altered how students recalled titration, it increased the instances of the dynamic hand-turning gesture for that condition. This represents another type of, and equally important, “knowledge unit,” one that is just as important as a verbal recall unit.

4.4.1 Epistemic network analysis (ENA) on knowledge units (both idea and gesture)

This study includes a type of multimodal analysis that can accommodate two types of knowledge units, both the verbal idea units and physical gesture units. Figure 9 helps visualize how the average projection of the high embodied group is situated on a different location along one axis. When the group graphs are subtracted, certain saliences emerge. Specifically, students in the control group used far fewer stopcock turning gestures (G3). The high embodied experimental group is expressing their comprehension of titration by also using “nonverbal knowledge”; if gesture is considered co-thought then more knowledge is being reported by the experimental group when recall is coded in a multimodal manner. Inspecting the strongest links in the subtraction ENA of Figure 9C, we see the highest values in the experimental group represent a triangle of I7 (Mention of Equipment) and G3 (HandTwist Gesture) and I9 (Terminate, the color stops changing). The G3 (HandTwist Gesture) played an essential role as a node in connecting the ideas and gestures in the experimental group, but the G3 gesture is very weakly linked for the control group. These results corroborated that the haptic MR device that aligned with a real experiment’s operation enhanced students’ knowledge recall on that titration operation. Specifically, students produced more post-intervention knowledge units when they manipulated the 3D-printed burette with a stopcock, a device that was designed to mimic a real titration device. These findings imply that designing for gestural congruency in a virtual environment helps to consolidate students’ knowledge both on the experimental operation and the principle.

4.5 RQ5: device preference

In the posttest phase, all participants had access to *both* interface devices and performed two titration tasks using both keyboard and 3D-printed burette. On a final survey question asking about preference for device, significantly more participants from *both* conditions preferred to use the 3D-printed haptic device. Participants also reported more interest and desire to use the high embodied and hands-on MR device. Thus, the 3D-printed haptic device was easy-to-use and preferred by a majority. Students typed in open-ended responses three of the more interesting ones are listed

here: 1) “Using this device was far more interactive and made me feel engaged. I’ve done virtual experiments before and always felt robbed of the experience but with this interactive burette I felt like I was actually in control of the experiment and that I was actually in a lab once again”, 2) “I preferred the device because it was more fun to play with there was a physical aspect to the lab”, and 3) “I preferred the device that I could physically touch because it kept me more interested and interested with the experiment.” This suggests that such devices would be relatively popular if available to the public.

4.6 Limitations and future studies

This study started during the beginning of the COVID-19 pandemic and so there was not an opportunity to compare the 3D-printed haptic burette condition with a real glassware lab condition. That is ultimately what should be done. One hypothesis is that using the 3D-printed haptic burette before getting into a real lab would have a positive effect for learning and efficacy, in line with the “preparation for learning” literature. That literature shows that learning and transfer are accelerated by what has transpired *before* accessing a real, or a more complex, condition (Schwartz and Martin, 2004; Chin et al., 2010).

In this titration experiment, our participant demographic consisted of college-age students. Given that the titration experiments are taught primarily in high school curriculum, we would like to be able to also conduct our titration experiment in a high school setting. There may be developmental effects that are of interest.

We did not see significant main effects between the conditions on the general chemistry knowledge measures and this may be because both conditions may have been somewhat agentic and active. Users in the control condition were able to control the burette drip speed by the duration of the key press, thus they had some “agency” in making choices and controlling digitized content. We make a somewhat finer distinction between low and high embodied in this experiment. Physically *turning* the burette to the left and right direction as occurs with real lab glassware is considered more embodied and *gesturally congruent*, compared to tapping keyboard arrow buttons (see Taxonomy for Embodiment in Education, Johnson-Glenberg, 2018). To unearth significant conditional differences on the traditional declarative knowledge assessments, we may need to include a *very* low embodied condition, perhaps one where participants passively watch videos of titrations being performed. Although, in truth, this lab is not very interested in setting up a passive control condition that has already been shown to result in lower amounts of chemistry knowledge being gained, this was seen in the Wolski and Jagodziński (2019) Kinect experiment described earlier.

More refined measures should be used to gather cognitive load metrics beyond the usual self-report. We did not include those in this study because we did not want to interrupt the learning phase with the usual load and effort questions. Future studies might include relatively non-invasive biometrics like pupil dilation (Robison and Brewer, 2020) and EEG to assess cognitive load.

We end by acknowledging that optimizing user interfaces for learning is a difficult task, even for simpler 2D computer screens (Clark and Mayer, 2016). The interface design space for AR and MR

is probably even *more* complex (Li et al., 2022; Radu et al., 2023). Designing for learning in XR spaces is an active field of research and an accepted design canon has yet to emerge. When working with haptic devices and the hands, designers also need to be aware of hand proximity effects (Brucker et al., 2021). We note that the 3D-printed burette was slightly more removed from the visual workspace than the connected laptop keyboard. Designers should also include universal design with sensitivity for those with disabilities and paralysis.

The goal is for this study to add some actionable information to the design field for STEM education using haptic MR devices. There were several interesting and significant ATIs that supported the same finding. That is, when learners came to the virtualized titration experience with *lower* prior knowledge and they were assigned to the haptic MR condition, those learners *performed worse* than the lower prior knowledge students assigned to the more traditional interface. The crossover interaction supports the contention that higher prior knowledge learners perform better in the haptic MR condition. Such studies help to inform educators and creators of online science courses that contain laboratory components.

5 Conclusion

Optimal interface design for learning modules using Mixed Reality (MR) is a complex and evolving space. This study compared a 2D desktop condition with a traditional keyboard input device to a higher embodied, passive haptic 3D-printed device. 3D-printed devices may serve as cost-effective bridges between the real glassware experience in chemistry and the purely digitized desktop experience. The use of cost-effective 3D-printed devices serves to reify real world skills and should be especially useful in rural or low SES communities that cannot afford expensive equipment. Results showed that students with more prior science knowledge and higher science identity scores performed better when they were in the tangible haptic experimental condition. The majority of students preferred using the 3D-printed haptic burette compared to the usual keyboard.

This study reveals some intriguing aptitude by treatment interactions based on a student's prior science knowledge scores. One recommendation for instructors may be to take into account the students' knowledge profiles and assign the lower prior knowledge students first to a more common device (one example is a keyboard) for scaffolding or a type of gradual exposure to the content. After those students demonstrate some science knowledge gains, they could then be switched to the more novel, real world and gesturally congruent device (e.g., the 3D-printed haptic burette). It may be the case that adding a haptic channel to the learning signal (which is usually only visual and auditory) may be overloading for lower prior knowledge students. These types of studies help to inform educators and creators of online science courses on how to make decisions regarding the delivery of laboratory instruction courses. They can make more informed decisions about *when and which types of students* should be moved over to a mixed reality interface. This may inform when instructors mail out 3D-printed kits to students, or when they invite students to conduct real experiments in a campus lab.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Arizona State University IRB. The participants provided their written informed consent to participate in the study.

Author contributions

MJ-G contributed to conception and design of the study. Co-wrote content, trained the experimenters, wrote the first draft of the manuscript, and approved the submitted version. CY gathered tests, ran participants, oversaw experiments, and trained experimenters wrote the Methods section, ran regression statistics, FL co-designed the content, and built the burette, CA wrote knowledge tests, co-designed the content, contributed to conception, YB contributed to concept of idea units, scored all units and aided in running ENA analyses, SY conceived of ENA analyses, ran all ENA analyses aided with regressions, RL contributed to conception, wrote the grant that secured funding, oversaw creation of the burette, approved submitted version. All authors contributed to the article and approved the submitted version.

Funding

Research supported by National Science Foundation grant number 1917912 to the first and last authors.

Acknowledgments

Dr. Pamela Marks, Dr. Mehmet Kosa, Augustin Bennett, Abigail Camisso, Aarushi Bharti, Rebecca Root, Ali Abbas, Kiana Guarino, Kambria Armstrong, Hannah Bartolomea, and Byron Lahey.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Aldosari, S. S., and Marocco, D. (2015). "Using haptic Technology for education in chemistry," in *Paper presented at the 2015 fifth international conference on e-learning (econf)* 18-20 Oct 2015.

Argüello, J. M., and Dempski, R. E. (2020). Fast, simple, student generated augmented reality approach for protein visualization in the classroom and home study. *J. Chem. Educ.* 97 (8), 2327–2331. doi:10.1021/acs.jchemed.0c00323

Barmoutis, A., Faris, R., Garcia, L., Gruber, L., Li, J., Peralta, F., et al. (2020a). "Assessing the role of virtual reality with passive haptics in music conductor education: A pilot study," in *Proceedings of the 2020 human-computer interaction international conference*. Editors J. Y. C. Chen and G. Fragomeni (LNCS), 275–285.

Barmoutis, A., Faris, R., Garcia, S., Li, J., Philocete, J., Puthusseril, J., et al. (2020b). "Virtual kayaking: A study on the effect of low-cost passive haptics on the user experience while exercising," in *Proceedings of the 2020 HCI international conference C*.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639

Bege, M., Ninaus, M., Schneider, S., Nebel, S., Schlemmel, J., Weidenmüller, J., et al. (2020). Investigating the effects of beat and deictic gestures of a lecturer in educational videos. *Comput. Educ.* 156, 103955. doi:10.1016/j.compedu.2020.103955

Born, F., Masuch, M., and Hahn, A. (2020). "Ghost sweeper: Using a heavy passive haptic controller to enhance a room-scale VR exergame," in *2020 IEEE conference on games (CoG)*, 221–228.

Bozbib, E., Baily, G., Haliyo, S., and Frey, P. (2021). 'Can I touch this?': Survey of virtual reality interactions via haptic solutions. *CoRR*, abs/2101.11278. Retrieved from <https://arxiv.org/abs/2101.11278>.

Bozgeyikli, E., and Bozgeyikli, L. L. (2021). "Evaluating object manipulation interaction techniques in mixed reality: Tangible user interfaces and gesture," in *Paper presented at the 2021 IEEE virtual reality and 3D user interfaces (VR)*. (27 March-1 April 2021).

Broaders, S. C., Cook, S. W., Mitchell, Z., and Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *J. Exp. Psychol. General* 136 (4), 539–550. doi:10.1037/0096-3445.136.4.539

Brucker, B., Brömmle, R., Ehrmann, A., Edelmann, J., and Gerjets, P. (2021). Touching digital objects directly on multi-touch devices fosters learning about visual contents. *Comput. Hum. Behav.* 119, 106708. doi:10.1016/j.chb.2021.106708

Chan, P., Van Gerven, T., Dubois, J.-L., and Bernaerts, K. (2021). Virtual chemical laboratories: A systematic literature review of research, technologies and instructional design. *Comput. Educ. Open* 2, 100053. doi:10.1016/j.caeo.2021.100053

Chang, H.-Y., and Tsai, C.-C. (2023). Epistemic network analysis of students' drawings to investigate their conceptions of science learning with Technology. *J. Sci. Educ. Technol.* 32, 0123456789. doi:10.1007/s10956-022-10026-9

Chin, D. B., Dohmen, I. M., Cheng, B. H., Oppezzo, M. A., Chase, C. C., and Schwartz, D. L. (2010). Preparing students for future learning with Teachable Agents. *Educ. Technol. Res. Dev.* 58 (6), 649–669. doi:10.1007/s11423-010-9154-5

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.

Cook, S. W., Mitchell, Z., and Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition* 106 (2), 1047–1058. doi:10.1016/j.cognition.2007.04.010

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information Technology. *MIS Q.* 13 (3), 319. doi:10.2307/249008

Drigas, A., and Gkeka, E. (2016). Montessori method and ICTs. *Int. J. Eng. Sci.* 4 (1), 25–30.

Dunleavy, M. (2014). Design principles for augmented reality learning. *TechTrends* 58 (1), 28–34. doi:10.1007/s11528-013-0717-2

Dunnagan, C. L., Dannerberg, D. A., Cuales, M. P., Earnest, A. D., Gurnsey, R. M., and Gallardo-Williams, M. T. (2020). Production and evaluation of a realistic immersive virtual reality organic chemistry laboratory experience: Infrared spectroscopy. *J. Chem. Educ.* 97 (1), 258–262. doi:10.1021/acs.jchemed.9b00705

Engelkamp, J. (2001). "Action memory: A system oriented approach," in *Memory for action: A distinct form of episodic memory? Counter points, cognition memory and language*. Editors H. D. Zimmer and R. L. Cohen (Oxford University Press), 49–96.

Engelkamp, J., and Zimmer, H. D. (1985). Motor programs and their relation to semantic memory. *Ger. J. Psychol.* 9, 239–254.

Flood, V. J., Amar, F. G., Nemirovsky, R., Harrer, B. W., Bruce, M., and Wittmann, M. C. (2015). Paying attention to gesture when students talk chemistry: Interactional resources for responsive teaching. *J. Chem. Educ.* 92 (1), 11–22. doi:10.1021/ed400477b

Fuhrman, O., Eckerling, A., Friedmann, N., Tarrasch, R., and Raz, G. (2020). The moving learner: Object manipulation in virtual reality improves vocabulary learning. *J. Comput. Assisted Learn.* 37 (3), jcal12515–683. doi:10.1111/jcal.12515

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdiscip. Rev. Cognitive Sci.* 1 (4), 586–596. doi:10.1002/wcs.55

Glenberg, A. M. (2008). "Embodiment for education," in *Handbook of cognitive science: An embodied approach*. Editors P. Calvo and A. Gomila (Amsterdam: Elsevier), 355–372.

Goldin-Meadow, S. (2011). Learning through gesture. *WIREs Cognitive Sci.* 2, 595–607. doi:10.1002/wcs.132

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychol. Sci.* 12 (6), 516–522. doi:10.1111/1467-9280.00395

Goldin-Meadow, S. (2014). Widening the lens: What the manual modality reveals about language, learning and cognition. *Philosophical Trans. R. Soc. Biol. Sci.* 369, 20130295. doi:10.1098/rstb.2013.0295

Han, D. T., Suhail, M., and Ragan, E. D. (2018). Evaluating remapped physical reach for hand interactions with passive haptics in virtual reality. *IEEE Trans. Vis. Comput. Graph.* 24 (4), 1467–1476. doi:10.1109/TVCG.2018.2794659

Hostetter, A., and Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bull. Rev.* 26 (3), 721–752. doi:10.3758/s13423-018-1548-0

Huwer, J., and Seibert, J. (2018). A new way to discover the chemistry laboratory: The augmented reality laboratory-license. *World J. Chem. Educ.* 6 (3), 124–128. doi:10.12691/wjce-6-3-4

Insko, B. E. (2001). *Passive haptics significantly enhances virtual environments*. The University of North Carolina at Chapel Hill.

Jiménez, Z. A. (2019). *Teaching and learning chemistry via augmented and immersive virtual reality*. Technology Integration in Chemistry Education and Research, 31–52. doi:10.1021/bk-2019-1318.ch003

Johnson-Glenberg, M. C., Birchfield, D., Koziupa, T., and Tolentino, L. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *J. Educ. Psychol.* 106 (1), 86–104. doi:10.1037/a0034008

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Front. Robotics AI* 5 (81), 81. doi:10.3389/frobt.2018.00081

Johnson-Glenberg, M. C. (2019). "The necessary nine: design principles for embodied VR and active stem education," in *Learning in a Digital World: Perspective on Interactive Technologies for Formal and Informal Education*. Editors P. Diaz, A. Ioannou, K. K. Bhagat, and J. M. Spector (Singapore: Springer Singapore), 83–112.

Kintsch, W. (1974). *The representation of meaning in memory*. Lawrence Erlbaum.

Kong, C. I., Welfare, J. G., Shenouda, H., Sanchez-Felix, O. R., Floyd, J. B., Hubal, R. C., et al. (2022). Virtually bridging the safety gap between the lecture Hall and the research laboratory. *J. Chem. Educ.* 99 (5), 1982–1989. doi:10.1021/acs.jchemed.2c00096

Kontra, C., Lyons, D., Fischer, S., and Beilock, S. L. (2015). Physical experience enhances science learning. *Psychol. Sci.* 26, 737–749.

Lages, W. S., and Bowman, D. A. (2019). "Walking with adaptive augmented reality workspaces: Design and usage patterns," in *Paper presented at the proceedings of the 24th international conference on intelligent user interfaces* (Marina del Ray, California). doi:10.1145/3301275.3302278

Laviola, J. J., Kruifff, E., McMahan, R. P., Bowman, D. A., and Poupyrev, I. (2017). *3D user interfaces: Theory and practice*. 2nd ed. Boston: Addison-Wesley.

Li, Y., Tahmid, I. A., Lu, F., and Bowman, D. A. (2022). "Evaluation of pointing ray techniques for distant object referencing in model-free outdoor collaborative augmented reality," in *IEEE Transactions on Visualization and Computer Graphics* 28 (11), 3896–3906. doi:10.1109/TVCG.2022.3203094

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frvir.2023.1047833/full#supplementary-material>

Lindgren, R., and Johnson-Glenberg, M. C. (2013). Emboldened by embodiment: Six precepts regarding the future of embodied learning and mixed reality technologies. *Educ. Res.* 42 (8), 445–452. doi:10.3102/0013189X13511661

Lowry, P., Gaskin, J., Twyman, N., and Hammer, B., and Oklahoma State University, and Louisiana Technical University (2013). Taking “fun and games” seriously: Proposing the hedonic-motivation system adoption model (HMSAM). *J. Assoc. Inf. Syst.* 14 (11), 617–671. doi:10.17705/1jais.00347

Macedonia, M. (2019). Embodied learning: why at school the mind needs the body. *Front. Psychol.* 10. doi:10.3389/fpsyg.2019.02098

Martinez, M., Morimoto, T., Taylor, A., Barron, A., Pultorak, J., Wang, J., et al. (2016). “3-D printed haptic devices for educational applications,” in *2016 IEEE haptics symposium (HAPTICS)*, 126–133.

Mayer, R. E. (2009). *Multimedia learning*. 2nd ed. New York: Cambridge University Press.

McNeil, N. M., Uttal, D. H., Jarvin, L., and Sternberg, R. J. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learn. Instr.* 19, 171–184. doi:10.1016/j.learninstruc.2008.03.005

McNeill, D. (2008). “Gesture and thought,” in *Gesture and thought* (University of Chicago press).

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Milgram, P., and Kishino, A. F. (1994). Taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* E77-D (12), 1321–1329.

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. *J. Chem. Educ.* 96 (3), 593–596. doi:10.1021/acs.jchemed.8b00794

Nathan, M. J., and Walkington, C. (2017). Grounded and embodied mathematical cognition: Promoting mathematical insight and proof using action and language. *Cognitive Res. Princ. Implic.* 2 (1), 9. doi:10.1186/s41235-016-0040-5

Radu, I., and Schneider, B. (2019). “What can we learn from augmented reality (AR)? Benefits and drawbacks of AR for inquiry-based learning of physics,” in *Paper presented at the proceedings of the 2019 CHI conference on human factors in computing systems*.

Radu, I., Yuan, J., Huang, X., and Schneider, B. (2023). Charting opportunities and guidelines for augmented reality in makerspaces through prototyping and co-design research. *Comput. Educ. X Real.* 2, 100008. doi:10.1016/j.cexr.2023.100008

Robison, M. K., and Brewer, G. A. (2020). Individual differences in working memory capacity and the regulation of arousal. *Atten. Percept. Psychophys.* 82 (7), 3273–3290. doi:10.3758/s13414-020-02077-0

Schaufeli, W. B., Bakker, A. B., and Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educ. Psychol. Meas.* 66 (4), 701–716. doi:10.1177/0013164405282471

Schwartz, D. L., and Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition Instr.* 22 (2), 129–184. doi:10.1207/s1532690xci2202_1

Segal, A., Tversky, B., and Black, J. (2014). Conceptually congruent actions can promote thought. *J. Appl. Res. Mem. Cognition* 3 (3), 124–130. doi:10.1016/j.jarmac.2014.06.004

Shaffer, D. W., Collier, W., and Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *J. Learn. Anal.* 3 (3), 9–45. doi:10.18608/jla.2016.33.3

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., et al. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. *Int. J. Learn. Media* 1 (2), 33–53. doi:10.1162/ijlm.2009.0013

Shaffer, D. W., and Ruis, A. R. (2017). “Epistemic network analysis: A worked example of theory-based learning analytics,” in *Handbook of learning analytics* (Society for Learning Analytics Research (SoLAR), 175–187. doi:10.18608/bla.17.015

Sheppard, K. (2006). High school students’ understanding of titrations and related acid-base phenomena. *Chem. Educ. Res. Pract.* 7, 32–45. doi:10.1039/B5RP90014J

Stets, J. E., Brenner, P. S., Burke, P. J., and Serpe, R. T. (2017). The science identity and entering a science occupation. *Soc. Sci. Res.* 64, 1–14. doi:10.1016/j.ssresearch.2016.10.016

Tee, N. Y. K., Gan, H. S., Li, J., Cheong, B. H.-P., Tan, H. Y., Liew, O. W., et al. (2018). Developing and demonstrating an augmented reality colorimetric titration tool. *J. Chem. Educ.* 95 (3), 393–399. doi:10.1021/acs.jchemed.7b00618

Ulrich, M., Evans, B., Liu, F. W., Johnson-Glenberg, M., and LiKamWa, R. (2021). “AR-some chemistry models: Interactable 3D molecules through augmented reality,” in *Adjunct publication of the 23rd international conference on mobile human-computer interaction* (France: Toulouse & Virtual). doi:10.1145/3447527.3474874

Varela, F., Thompson, E., and Rosch, E. (1991). *The embodied mind: cognitive science and human experience*. Cambridge, MA: MIT Press.

Walkington, C., Nathan, M. J., Huang, W., Hunnicutt, J., and Washington, J. (2023). *Multimodal analysis of interaction data from embodied education technologies*. Educational Technology Research & Development.

Wasserman, S., and Faust, K. (1994). *Social networks analysis: Methods and applications*. Cambridge: Cambridge University Press.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bull. Rev.* 9 (4), 625–636. doi:10.3758/bf03196322

Wolski, R., and Jagodziński, P. (2019). Virtual laboratory—using a hand movement recognition system to improve the quality of chemical education. *Br. J. Educ. Technol.* 50 (1), 218–231. doi:10.1111/bjet.12563

Wu, L., Liu, Q., Mao, G., and Zhang, S. (2020). Using epistemic network analysis and self-reported reflections to explore students’ metacognition differences in collaborative learning. *Learn. Individ. Differ.* 82, 101913. doi:10.1016/j.lindif.2020.101913

Yang, Z., and Weng, D. (2016). “Passive haptics based MR system for geography teaching,” in *Proceedings of the 15th ACM SIGGRAPH conference on virtual-reality continuum and its applications in industry - volume 1* (China: Presented at the Zhuhai), 23–29. doi:10.1145/3013971.3013995

Zacharia, Z. C. (2015). Examining whether touch sensory feedback is necessary for science learning through experimentation: A literature review of two different lines of research across K-16. *Educ. Res. Rev.* 16, 116–137. doi:10.1016/j.edurev.2015.10.001

Zhang, S., Chen, J., Wen, Y., Chen, H., Gao, Q., and Wang, Q. (2021). Capturing regulatory patterns in online collaborative learning: A network analytic approach. *Int. J. Computer-Supported Collab. Learn.* 16 (1), 37–66. doi:10.1007/s11412-021-09339-5