

Collaborative Embodied Learning in Mixed Reality Motion-Capture Environments: Two Science Studies

Mina C. Johnson-Glenberg, David A. Birchfield, Lisa Tolentino, and Tatyana Koziupa
Arizona State University

These 2 studies investigate the extent to which an Embodied Mixed Reality Learning Environment (EMRELE) can enhance science learning compared to regular classroom instruction. Mixed reality means that physical tangible and digital components were present. The content for the EMRELE required that students map abstract concepts and relations onto their gestures and movements so that the concepts would become grounded in embodied action. The studies compare an immersive, highly interactive learning platform that uses a motion-capture system to track students' gestures and locomotion as they kinesthetically learn with a quality classroom experience (teacher and content were held constant). Two science studies are presented: chemistry titration and disease transmission. In the counterbalanced design 1 group received the EMRELE intervention, while the other group received regular instruction; after 3 days and a midtest, the interventions switched. Each study lasted for 6 days total, with 3 test points: pretest, midtest, and posttest. Analyses revealed that placement in the embodied EMRELE condition consistently led to greater learning gains (effect sizes ranged from 0.53 to 1.93), compared to regular instruction (effect sizes ranged from 0.09 to 0.37). Order of intervention did not affect the final outcomes at posttest. These results are discussed in relation to a new taxonomy of embodiment in educational settings. We hypothesize that the positive results are due to the embodiment designed into the lessons and the high degree of collaboration engendered by the co-located EMRELE.

Keywords: collaboration, embodied learning, virtual reality, media in education, embodiment, science education

Supplemental materials: <http://dx.doi.org/10.1037/a0034008.supp>

There was a general expectation with the wiring and technological immersion of American schools that great strides would be made in

math, science, and reading skills. Several studies have demonstrated that merely allowing students access to educational software and technology does not in itself result in significant learning gains. A U.S. Department of Education-sponsored study on the effects of several popular commercialized reading and math software programs found that test scores did not significantly differ depending on software treatment versus regular instruction control classroom (Dynarski et al., 2007). Another result, from a 4-year technology immersion study in Texas with 42 schools (TCER; Texas Center for Educational Research, 2009), demonstrated that the technology effects on student learning were positive overall, but there was not a sustained significant effect by the fourth year. These studies were performed on commercial programs that, at the time, used low levels of interactivity. It should be noted that computer aided instruction created by research groups was, at the time, demonstrating significant gains in learning (e.g., interactive intelligent tutoring systems using smaller grain sizes of feedback (van Lehn, 2011) and strategy training systems (Graesser, McNamara, & van Lehn, 2005; Johnson-Glenberg, 2005, 2007). However, many of these research-based systems have yet to be commercialized.

There may be several reasons why exposure to commercialized educational technology has not significantly improved education in America. Many studies (including the TCER one) cite the lack of professional training and infrastructure support, but we posit that another hindrance to sustained learning gains might be the *design* of the content in addition to the *delivery environment* itself. All educational technology is not the same.

This article was published Online First September 16, 2013.

Mina C. Johnson-Glenberg, Learning Sciences Institute and Department of Psychology, Arizona State University; David A. Birchfield, Lisa Tolentino, and Tatyana Koziupa, School of Arts, Media and Engineering, Arizona State University.

David A. Birchfield is now at *SMALLab* Learning, LLC, Los Angeles, California.

Mina C. Johnson-Glenberg and David A. Birchfield are co-founders of the University Spinout Company. This research was supported in part by Intel, the MacArthur Foundation (under a grant titled "Gaming *SMALLab*"), the National Science Foundation CISE Infrastructure Grant 0403428 and IGERT Grant 050464. We would also like to thank the *SMALLab* group at School of Arts, Media and Engineering at Arizona State University; Colleen Megowan-Romanowicz; Christobal Martinez; Sibel Uysal; and the teachers, administrators, and students at Coronado High School who worked with us.

Please see the following links for further information regarding the studies in this article: For Chemistry Experiment 1 videos created for teacher training by lead designer: <http://www.youtube.com/watch?v=EwEF0L3ROb0&feature=c4-overview-vl&list=PLD4761BA48618FF7B>; For Disease Transmission Experiment 2 videos created for teacher training by teacher in the study: <https://www.youtube.com/watch?v=As2zGw2bj7g&list=PL83801B5EFACB507D>.

Correspondence concerning this article should be addressed to Mina C. Johnson-Glenberg, Learning Sciences Institute, 1000 Forrest Mall, ASU, Tempe, AZ 85287. E-mail: mina.johnson@asu.edu

Traditionally desktop computers have been the method for delivery. The first wave of commercialized educational software available to schools around 2000 resembled little more than scanned textbooks. The software was very linear and, in general did not take advantage the computer's interactive capacity for adaptivity and integration of user-constructed content. Today, the prevalent technology model in schools is still one-student-to-one-screen, this model does not take advantage of rich sociocollaborative interactions that are rife in school settings. There is evidence that many technology-based learning environments are most effective when coupled with face-to-face interactions (Asllani, Etkin, & Somasundar, 2008). Constructive learning environments, media-rich technology, and collaborative learning can co-occur. When the digitally delivered content is properly designed and delivered in an environment that facilitates collaboration, we posit that impressive learning gains can result. In this article, we explore the types of learning gains that can be expected when students learn in an Embodied Mixed Reality Learning Environment (EMRELE). Mixed reality (Milgram & Kishino, 1994) is the term used to describe computer-supported environments where both physical objects (e.g., a tracking wand, the whiteboards students use to make notes) and virtual objects (e.g., the projected digital content) are used. Two experiments are presented that address several unanswered questions associated with such environments: (a) How do learning gains in EMRELEs compare to those seen in regular instruction classrooms when teacher and content are held constant, and (b) does the order of intervention matter?

There is consensus among educational researchers when using digital learning strategies that teachers should be moving toward more student-centered and more collaborative techniques (Smaldino, Lowther, & Russell, 2008). Being collaborative learners requires the learners to be explicit; Crook (1996) suggested that these processes of articulation, and the ensuing discourse conflicts, lead to a co-construction of ideas that is often better than constructions created by a single learner. Peers in a problem-solving situation are often obligated to make justifications and negotiations. Eventually peers should be able to converge on a common goal of shared understanding. This article presents research on a direct comparison between regular classroom instruction with co-location and group activities to a technology-supported learning environment with co-located collaborative group activities while students are immersed in an embodied mixed reality environment.

Example of an EMRELE

The Situated Multimedia Arts Learning Lab (*SMALLab*; Johnson-Glenberg, Birchfield, & Uysal, 2009) is an example of an EMRELE that uses motion-capture and a highly collaborative pedagogy. *SMALLab* is an educational platform that engages the major modalities (i.e., the sense systems including visual, auditory, and kinesthetic) that humans use to learn. The platform is kinesthetic, scalable, and easy to enter/exit. *SMALLab* uses 12 infrared OPTITRACK motion tracking cameras that send information to a computer about where a student holding a tracked object is in a floor-projected environment. The floor space is 15×15 feet (4.572×4.572 m), and the tracked space extends approximately eight feet high. Students step into the active space and grab a "wand" (a rigid body trackable object) that allows the physical body to now function like a 3D cursor in the interactive space.

The environment also allows for multiple students (up to four) to be tracked simultaneously. With turn-taking, entire classrooms with 30 students are able to physically experience a learning scenario in a typical class period. Students outside the active space sit around the open periphery and collaborate by discussion and whiteboard activities with each other, and with the active students. Figure 1A represents a schematic of the system; Figure 1B shows students seated around the *SMALLab* floor projection engaged in an earlier titration scenario.

When *SMALLab* is constructed in a classroom the cameras hang from the ceiling and the truss system is not needed. Figure 1 shows an earlier version of the titration scenario that used glowing track balls (we now use more ergonomic wands that the system can track with millimeter precision). The trackable objects are used to select virtual molecules from the edge of the floor and the system uses the height of the object to serve as the release mechanism. That is, when the tracked object drops below a threshold the virtual molecule is "released" into the projected flask in the middle of the floor. Teachers use remote controls to control the flow of the lesson and select which dynamic graphical and/or sonic media the students interact with.

Well-designed mixed reality can allow students to perceive feedback via multiple modalities and in a group setting. In our centripetal force physics scenario students are able to *hear* immediate sonic feedback in proportional pitch to the acceleration of a swinging physical tracked object, as they *see* the virtual motion map with vectors plotted on the floor in real time; in addition, they *feel* how

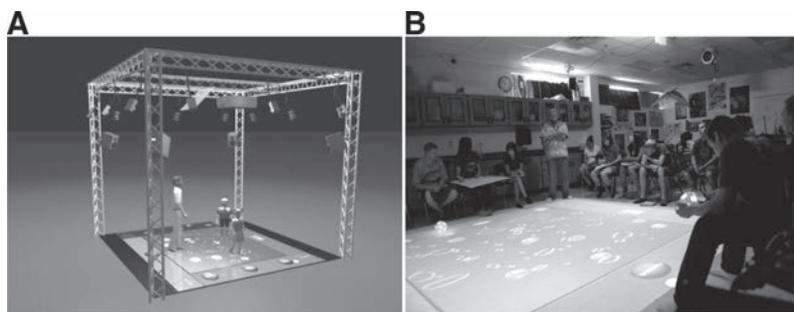


Figure 1. A. Schematic for a free standing *SMALLab* learning environment. B. Students sitting around the embodied environment at a high school with cameras and projector embedded in the ceiling.

their body is affected by the centripetal force on the swinging object. The embodied environment allows the body to function as another feedback signal, and the timing of all feedback is designed to optimize learning in a nondisruptive manner (Shute, 2008).

SMALLab is similar to other immersive, mixed-reality environments such as CAVE environments (small areas with rear-projection screens for walls and a down-projection screen for the floor; Cruz-Niera, Sandin, & Defanti, 1993). Immersive, visually rich learning environments are becoming more affordable. Iowa State now displays its *C6* 10 feet \times 10 feet \times 10 feet (3.048 m \times 3.048 m \times 3.048 m) room where all four walls and the floor and ceiling are screens with back-projected stereoscopic images. *Elumenati* is a learning company that specializes in permanent or portable blow-up domes that simulate immersive “planetarium-like” experiences for small groups. *SMALLab* uses one large projection on the floor and is a rigid body motion-capture environment, as opposed to the laser tracking learning environment used by Lindgren and Moshell (2011). These last two EMRELE’s are physically open on all sides, thus allowing students to directly communicate with peers inside and outside of the active space. This feature is important because it facilitates collaboration and observational learning opportunities.

Previous Research

We have researched EMRELEs in several different content domains, including language arts (Hatton, Birchfield, & Megowan, 2008); science, technology, engineering, and mathematics (STEM) content (Birchfield & Megowan-Romanowicz, 2009; Tolentino et al., 2009), and special education (Savvides, Tolentino, Johnson-Glenberg, & Birchfield, 2010). This range suggests that embodied learning is not content dependent. One question we had about highly embodied learning was whether the order of intervention would affect final learning gains.

Order of Intervention

The counterbalanced design allows us to ask, “If teachers had time to teach using both an EMRELE and regular instruction, would it matter which method came first?” A previous geology study that lasted for 6 days examined student learning gains regarding the Earth’s complex, dynamic strata (Birchfield & Johnson-Glenberg, 2010; Birchfield & Megowan-Romanowicz, 2009). Students were randomly assigned to receive either the EMRELE or regular instruction first, after 3 days the interventions switched. Statistically significant learning gains were seen whenever the students were in the technology-supported EMRELE intervention. In the regular instruction intervention (with teacher held constant), students created hands-on paper timelines and discussed the dynamics of geology in small groups. It was a collaborative and appropriate control that resulted in learning gains; however, those gains were not statistically significant. At the final test point (Time 3) the group that received regular instruction *then* *SMALLab* demonstrated significantly higher test scores, compared to the group that received *SMALLab* *then* regular instruction. One hypothesis is that the regular instruction laid the foundation for the more complex learning that would occur later in the *SMALLab* (EMRELE) intervention. Our prediction for the studies presented here is also that the groups that receive the

SMALLab interventions second (after regular instruction) will demonstrate the greatest overall learning gains by final posttest.

Reasons for Learning Gains

We propose two primary reasons for the consistently higher gains seen in previous EMRELE studies: embodiment and collaboration. In the discussion section, we explore the contribution of novelty.

Embodiment

Embodiment is not a “passing fad” in psychology (Newman, 2008). Piaget (1952) was an early proponent that sensorimotor activity aids in constructing knowledge and that bodily actions are not separate from, nor solely downstream from, the mind. In our context, manipulatives are considered hand-held tangibles that can be used to manipulate digital objects for learning. This new, mixed reality environment blurs the distinction often made between hands-on science activities versus digital activities (lessons on computer screens). As Klahr, Triona, and William (2007) revealed, there are nuanced benefits associated with each condition. We posit that an environment that affords both may be optimal.

Cognitive processes have “deep roots in sensorimotor processing” (p. 625) and come from the body’s interactions with its physical environment (Wilson, 2002). Multiple research areas now support the tenet that embodiment is an underpinning of cognition. The various domains include (but are not limited to) neuroscience (Rizzolatti & Craighero, 2004), cognitive psychology (Barsalou, 2008; Glenberg, 2010; Glenberg & Kaschak, 2002), math (Lakoff & Nunez, 2000), gesture (Goldin-Meadow, 2011; Hostetter & Alibali, 2008), expert acting (Noice & Noice, 2006, with the idea of “active experiencing”), and dance (Winters, 2008).

Glenberg (2010) contends that human cognition comes from developmental, embodied interactions with physical environments. The theory is that all thought, even the most abstract, is derived from physical embodiment. Pulvermüller and Fadiga’s (2010) review of fMRI experiments demonstrate that when reading words related to action, areas in the brain are activated in a *somatotopic* manner. For example, reading “lick” activates motor areas that control the mouth, whereas reading “pick” activates areas that control the hand. This activation is part of a parallel network representing “meaning” and shows that the mappings do not fade once stable comprehension is attained. Motoric codes are still activated during linguistic comprehension in adulthood. Glenberg, Sato, and Cattaneo (2008) demonstrated that when motor systems are adapted, comprehension of sentences implying the use of those areas is affected. For example, after adapting to an *away from* the body action, participants are slower to accept as sensible sentences that describe action in the *away from* motion. This suggests that encoding and the motor system are coupled, and fatiguing or adapting the motor system can affect cognition.

A spate of studies in the domain of self-performed tasks (SPT) supports the contention that physical actions affect memory. A representative study compared three groups of participants: one that heard a list of unrelated action phrases (“lift the hat”), one that performed the action without the object, and one that performed the task with the object. The consistent finding was that the self-performing participants recalled more of the phrases than

those who merely heard the phrases (Engelkamp, 2001; Engelkamp & Zimmer, 1985).

Finally, increasing evidence in the study of gesture suggests that gestures facilitate speech about mental images (Hostetter & Alibali, 2008). Gesture may serve as a *cross-modal prime* to facilitate retrieval of mental or lexical items. If the physical movement primes (readies) other constructs (like language), then learning via movement may add an additional modality and prime for later recall of knowledge. Cook, Yip, and Goldin-Meadow (2010) found that gesturing during the learning of information improved recall whether the speaker chose to gesture spontaneously, or was instructed to gesture. They hypothesized that gesturing during encoding functioned like action in facilitating memory.

A more rigorous understanding of embodiment in education is needed and so we present a taxonomy of embodiment with current examples of technology and multimedia in education. At the low end of embodiment are desktop-based simulations that are passively viewed. There is a large literature that reviews the efficacy of multimedia visualizations (or “animated simulations”) and readers are directed to that literature (e.g., Plass, Homer, & Hayward, 2009; Rouet, Schnotz, & Lowe, 2008; Schnotz & Kirschner, 2007). Here we focus on distinctions at the higher end of the embodiment spectrum where gesture and grosser body movements can be incorporated. With the advent of cost-effective skeletal tracking (e.g., Microsoft *Kinect*) entering the education arena, there is a need to focus the conversation. This taxonomy is a first attempt to partition novel motion-sensing and embodied learning environments into meaningful categories. Certainly more research needs to be done on the claims about sensorimotor/afferent neuronal activation and the perception of immersion when learning, our goal is to get the conversation started. We would like to explore which affordances are necessary components in efficacious embodied learning design. Is there a way to discretely rank these affordances?

Taxonomy for Embodied Learning

We propose three necessary components in a taxonomy for embodied learning: (a) amount of motoric engagement, (b) gestural congruency (i.e., how well-mapped the evoked gesture is to the content to be learned), and (c) perception of immersion. These three components occur on three continuous axes, but it is not helpful to keep them continuous, thus, we partition the taxonomy into four categories or degrees. The edges between the degrees should be considered fuzzy. The fourth degree is the highest.

Fourth degree = Includes locomotion which results in a high degree of sensori-motoric engagement; gestures are consistently designed to map to content being learned; and learner perceives environment as very immersive.

Third degree = No sustained locomotion, but whole body could still be engaged while in same area; some amount of gestural relevancy; learner perceives environment as immersive.

Second degree = Learner is generally seated, there is upper body movement; interfaces should be highly interactive, but gestural relevancy is not a given; with smaller display (mon-

itor or tablet) the learner does not perceive the environment as highly immersive.

First degree = Learner is generally seated, some upper body movement; primarily observes video/simulation—no gestural relevancy; with smaller display learner does not perceive environment as immersive.

Table 1 lists the degrees of embodiment and several examples of educational applications. Again, to be considered embodied in the highest or fourth degree the following three components must be strongly represented: (a) Motoric engagement. Full body motoric engagement is achieved through locomotion, so a technology that affords the learner the opportunity to ambulate will score highest on this. Campos et al. (2000) contended that locomotion is still important and relevant for the adult. The broad-based and context-specific psychological reorganizations set in place via toddler locomotion have powerful consequences, and after infancy “can be responsible for an enduring role in development by maintaining and updating existing skills” (Campos et al., 2000, p. 210). For the infant, locomotion is a “setting event, a control parameter, and a mobilizer that changes the intrapsychic states of the infant” (Campos et al., 2000, p. 150). Locomotion effects changes in social and emotional development, referential gestural communication, wariness of heights, the perception of self-motion, distance perception, spatial search, use of parallax information, variance with attentiveness, and spatial coding strategies. Locomotion in a learning environment affects the user’s *optic flow*, i.e., the continuously changing ambient optic array produced by a continuously moving point of observation (Gibson, 1979).

(b) The second component, gestural congruency mapping, falls from the first. The degree of motoric engagement can be highly correlated with the amount of gestures; however, we are most concerned with the congruency or relevance of the gesture as it maps to the content to be learned. The gestures should be linked to the content in a manner that reifies the learning construct. When this occurs we claim it has “gestural relevancy.” As an example, if we want to instruct in the parity rule for gear trains, and we are designing with the Microsoft skeletal-tracking *Kinect* sensor as the input device, we would not want to start the gear system moving with a *push hand forward* motion. We would want to start the system turning by having the students circle their hands in the direction of the first input gear (if clockwise, they would circle right hand around right shoulder joint to the right; Johnson-Glenberg, 2012). Gestural relevancy has also been called “congruent gestural conceptual mapping” by others (Segal, Black, & Tversky, 2010). Segal et al. (2010) found when students used an iPad touch surface with congruent gestures (tapping on each block to count) versus noncongruent gestures (tapping on a number), they made significantly fewer errors. With the advent of multi-touch screens, it is important that gesture research become more codified. The term “embodied” is in danger being overused to the point of meaninglessness. Yes, pinching the screen on multitouch surfaces makes the image smaller, but if that diminution is not central to the content (e.g., how a sponge reacts to pressure), then the gesture would score low on the gesture congruency scale, and the lesson would fall into the first or second degree of embodiment in the proposed taxonomy.

Table 1
Taxonomy for Four Degrees of Embodiment in Educational Technology

Variable	Varying degrees of sensorimotor activation, gestural congruency and immersion			
	4th degree Mixed-Reality: Motion Sensing with Locomotion	3rd degree Motion Sensing Interfaces and/or Large Display	2nd degree Small Screen— Interactive	1st degree Small Screen— Observational/Passive
Examples	Interactive Motion Capture Environments with large displays, e.g., <i>SMALLab</i>	Large displays (e.g., Interactive Whiteboards) could have motion sensing peripherals or not (Microsoft <i>Kinect</i>) Examples include, <i>SMART Tables</i> , Head Mounted Displays, ^a Flight Simulators	Desktop monitor or tablet designed with generative interactivity, e.g., PhET Simulations— <i>Forces and Motion 2.0</i> ^b	Desktop monitor or tablet without generative interactivity; this is simply viewing videos or simulations, e.g., National Geographic for schools, original Kahn Academy
Degrees based on (a) Amount sensorimotor engagement (b) Type of engagement—is gesture congruent to content? (c) Immersion—perceived	Whole body locomotion; gestures and tangible manipulations are highly congruent to content to be learned; immersive or semi-immersive	Could engage whole body, but generally in one place; must include gestural congruency but usually not w/tangible manipulative, immersive or semi-immersive	Stationary; should include gestural congruency with strong interactivity; not perceived as immersive due to small screen	Stationary; no gestural congruency; contains no interactivity beyond starting a simulation; not perceived as immersive (Observational learning of perceptual symbols allows it to be called embodied)
Sensorimotor engagement				
Supports full body movements and locomotion	X	(Skeletal tracking system like <i>Kinect</i> could support limited locomotion)		
Supports gestures	X	X	X (usually relies on smaller mouse-driven movements or small screen touch movements)	

^a Head-mounted Displays are now being used for educational purposes, e.g., in aviation and medical fields. Liu, Jenkins, Sanderson, Fabian, and Russell (2010) used a Nomad ND2000 to assess whether anesthesiologists could be aided in keeping their gaze on the patient, rather than on the anesthesia workstation. ^b Forces and Motions (<http://phet.colorado.edu/en/simulation/forces-and-motion>) requires the user to (among other options) draw an applied force vector with a mouse. The length of the user-generated vector maps to the magnitude of the force and the object is animated accordingly.

(c) The third component is immersion, and we refer to this more accurately as the *perception of immersion* because it is subjective and strongly contingent on the type and configuration of the content’s display. Although 360° wrap-around rooms exist, these are still expensive and not used in K–12 education where space is at a premium. Virtual reality (VR) goggles and head-mounted displays (HDMs) have not taken off either. We include mobile VR in our metrics because the expectation is that these will eventually become cost effective and highly mobile (see Google Glass, perhaps even shrinking to the size of a contact lens by 2030; Kaku, 2011). The VR community makes strong claims that “head-tracked, egocentric camera control provides a stronger sense of immersion, or ‘being there’ than a desktop display” (Pausch, Proffitt, & Williams, 1997, p. 16). For first and second degree embodied systems, we assume the content is on smaller, vertical computer screens (“desktops” in this article) or on typical 10.1-in. (25.654-cm) screen size tablets. Although tablets and smartphones allow the user to be mobile and the embedded accelerometers can be used for motion tracking, we have not yet seen randomized controlled studies on mobile devices that push on embodied learn-

ing. Using a tablet-sized screen for touch purposes limits the magnitude of the gesture.

In first and second degree lessons learners are often dealing with “simulations,” we use this term to describe graphic, visually compelling content that could be interactive (meaning that the user moves content or constructs new entities like concept maps—this last act would be called “constructive” and not interactive, as Chi, 2009, used the terms). To earn the second degree label, the input should have some gestural congruency, e.g., drawing a longer vector onscreen with the mouse makes the virtual car travel further. However, the input device for desktops is often a generic mouse interface, and this activates less of the sensorimotor system (drawing over a short distance via wrist movement) versus drawing vectors across the expanse of a large interactive whiteboard (IWB; walking several feet, using core, shoulder, arm, wrist, etc.). The magnitude of the afferent and efferent motor system cascade is important. We hold that for smaller, monoscopic devices the size of the display affects a user’s perception of immersion or feeling of presence (see extent of presence metaphor analogy in Milgram & Kishino, 1994). We do not yet know the boundaries of this sense

of immersion and how it interacts with embodied learning. Continuing the vector example, if a user swipes a finger with acceleration across a 10.1-in. screen tablet to demonstrate acceleration that is surely embodied, but the constraints of the input surface size assure that the motion is smaller. It has been created with less sensorimotor activation compared to a longer vector swipe created with a whole arm movement on an IWB. At this stage in the research, we place tablet-delivered lessons with some gestural congruency in the second degree. For the first degree of embodiment, the learning is always passive and observational regardless of display size, e.g., watching a simulation with little user input or interactivity.

In sum, our claim is that for a learning module to be considered embodied to the highest degree it should activate multiple afferent and efferent neuronal pathways in the learner's motor system, and these movements should be gesturally congruent to the content to be learned. This combination of body-based muscle engagement, brain-based gesture-as-semantics loop and immersive display should result in 4th degree learning. Removing the body-based (kinesthetic) component, will still result in learning; however, the resultant observational learning may not be as durable (i.e., prone to long term retention). We are now seeing evidence for significant knowledge retention differences when comparing lessons with varying degrees of embodiment (Johnson-Glenberg, Birchfield, Glenberg, Megowan-Romanowicz, & Savio-Ramos, 2013). As Black, Segal, Vitale, and Fadjo (2012) posited, "the richer the perceptual experience, and therefore the mental perceptual simulation acquired, the better the student learning and understanding" (p. 199).

Instructional design for new media should take advantage of collaboration and the many opportunities for observational learning in a classroom. This is not well facilitated by each child looking at a single monitor or tablet. The collaboration that comes out of immersive, open platform learning is emergent and often unscripted, so the opportunities for it must be explicitly designed into the curriculum. Our teachers often split students into groups and give them simple whiteboards and markers. The students are asked to make predictions, discourse on them, and then record them for verification after the scenario. Large screen and/or floor projection displays lend themselves to multi-student, co-located interactions.

Collaboration

The second causal factor for the learning gains seen in previous EMRELE studies may be the enhanced collaboration between the students. In general, collaborative learning in the classroom generates significantly higher achievement outcomes, higher level reasoning, better retention, improved motivation, and better social skills (D. W. Johnson & Johnson, 1991, 1984, 1989) than traditional didactics. Recent research on collaboration in small computer-mediated groups reveals that learning and transfer gains are mediated by the profiles (level of expertise) of the students within the group (Nihalani, Mayrath, & Robinson, 2011). The effects of collaboration in virtual and game-based worlds are complex, multi-causal and dynamic. As an example of the complexity during gameplay, group learning can be significantly affected by whether the group ends up with a leader or not (Wendel, Gutjahr, Goebel, & Steinmetz, R.,

2013). Wendel et al.'s (2013) design approach takes into account the requirements of "traditional single player games (fun, narration, immersion, graphics, sound), challenges of multi-player games (concurrent gaming, interaction) and Serious Game design (seamless inclusion of learning content, adaptation and personalization) . . . group goals, positive interdependence, and individual accountability" (p. 287). Although designing co-located, game-based content is challenging the payoffs can be considerable. A recent metaanalysis on serious games found that players in a group of two or more learned more content compared to players who played individually (Cohen's $d = 0.66$; Wouters, van Nimwegen, Oostendorp, & van der Spek, 2013).

Although there is now an impressive body of research on digitized, asynchronous collaborative learning (e.g., Weinberger & Fisher, 2006), prior research on collaborative learning in a digitized face-to-face (F2F) EMRELEs is harder to find. Birchfield and Megowan-Romanowicz (2009) coded four types of verbal utterances and compared frequency of utterances between regular instruction in a geology classroom to an EMRELE condition (teacher and content held constant). They found elevated levels of student-to-student comments and multi-turn student conversations in the EMRELE, compared to more teacher-to-student and student-to-teacher comments in regular instruction. The early collaboration research was conducted on small groups without computers (D. W. Johnson & Johnson, 1991), but we still find the categories of the F2F collaboration helpful for EMRELEs. There are four perspectives on why collaboration produces effective learning: (a) The *motivational perspective* on cooperative learning focuses primarily on the reward or goal structures under which students operate. From this perspective (e.g., R. T. Johnson & Johnson, 2009), cooperative incentive structures create a situation where group members attain their own personal goals if the group as a whole is successful. (b) The *social cohesion perspective* is related to the motivational viewpoint in that the effects of collaboration on achievement are mediated by the cohesiveness of the group. If students care about one another, they will want to help each other succeed. (c) The *cognitive developmentalist perspective* suggests that collaboration promotes growth because learners of similar ages are likely to be operating within each other's zones of proximal development (ZPD; Vygotsky, 1978). That is, some of the more advanced learners model in the group behaviors to which the more striving learners have not yet been exposed. This can provide a forum for discovery learning and encourage creative thinking with idea generation. (d) The *cognitive elaboration perspective* assumes that each learner in the group gets a chance to elaborate (explain) the material, or an aspect of the material, to one another. Several of these perspectives interact during the complex collaboration present in our EMRELE with the social cohesion and cognitive developmentalist perspectives being most strongly represented in our design.

The studies presented here address two timely questions in two different content domains. The first study focuses on chemistry content and asks (a) Do students learn more in an EMRELE environment compared to quality regular instruction? (b) Does the order of intervention affect final learning outcomes? The second study focuses on biology content.

Study 1: Chemistry

Method

Participants. The participants came from a large urban high school. Free or reduced lunch was available to 68% of the students. Student ethnic breakdown was as follows: Hispanic 48%, Caucasian 36%, African American 8%, American Indian 5%, and Asian 3%. Participants came from three 10th–11th-grade chemistry classrooms, and the same teacher taught all three classes. He had been a teacher for 9 years and had taught chemistry exclusively for the past 4 years. There were 69 students in the three classes; of those only 51 took all three tests, and only those students are included in the analyses. An equal proportion was deleted from each condition due to missing test points from absences during test times; Group 1 = 11/46, Group 2 = 7/23.

Design. Both Experiments 1 and 2 used a repeated-measures design with a counterbalanced control group. This is essentially an AB versus BA design with three test points. The EMRELE system with its extensive hardware was installed in one high school classroom and so students came to that classroom like a lab. The two interventions, *SMALLab* instruction or regular instruction, were administered in 3-day increments. Classes were assigned to receive either the *SMALLab* (the EMRELE style) intervention first or regular instruction first. On Day 4, of the 6-day long intervention, order switched. Thus on Day 4, the students who had been in *SMALLab* now received regular instruction and the students who had been in regular instruction received *SMALLab*. The teacher remained the same. By the end of Day 6, all students had experienced both types of instruction (see Table S1 in the online supplemental materials).

Because this study occurred at the end of the school year, the teacher was confident that his three classes were performing equivalently (this was later confirmed with the pretest analysis). Due to scheduling and logistical constraints regarding access to the *SMALLab* classroom (e.g., moving all desks around), the decision was made to place the two contiguous periods (i.e., Periods 2 and 3, which were back-to-back) in one condition, and the later class (Period 6) in the other condition. Thus, the classrooms were quasirandomly assigned to condition. On day one, all students took Test 1 (pretest). One class received 3 days of *SMALLab* instruction (Group 1, $n = 16$), and then 3 days of regular instruction. The other group (Group 2, $n = 35$) received 3 days of regular classroom instruction, followed by 3 days of *SMALLab*. These interventions ran concurrently, so that at the end of Day 3 all students took Test 2 (midtest). At the end of Day 6, all students took Test 3 (posttest). The test was invariant. Any effects due to familiarity with exact test form would be the same for both conditions; these also include test fatigue effects.

***SMALLab* Instruction.** A lesson plan can be found in the online supplemental materials. The chemistry titration scenario was co-designed by two teachers who have a combined 20 years of teaching experience and profess to use “inquiry-based teaching” (Llewellyn, 2005). The *SMALLab* scenario was modeled after a traditional chemistry titration lab. In typical titration labs, students use glassware to gradually add a known solution of acid or base to a known reactant of unknown molarity until the endpoint of reaction occurs. This end point is often signaled by the solution turning pink. Our goal was to match and *enhance* what happens in

the teacher’s hands-on science lab with the EMRELE’s multimedia and embodied capabilities. The media allowed the teacher or student to stop the molecular processes at any point for more in-depth discussion.

On the first day the instructor made sure that each student had a chance to explore how the tracking wand could be used to select and move virtual molecules in the space. He ascertained that all knew what the molecular symbols represented. Two students were active in the space at once; one controlled the acids, the other the bases. The remainder sat around the perimeter where they could see the floor projection, hear the audio feedback and offer hypotheses. A screenshot of the projected visual interface is shown in Figure 2. See the Author Note for links to the videos.

This central area in the projection represents a virtual “flask,” where the interactive titration process unfolds. Students grab molecules with the wand and then toss them into the virtual flask area. Surrounding the virtual flask are four panels. The top panel consists of three actionable (movable) *acid* molecules, and the bottom panel consists of three actionable *base* molecules. Along the left edge is one choice of indicator particle. Along the right panel is a pH level number that dynamically adjusts depending on the pH of the virtual flask. To select a molecule, the student holds the wand over a molecule image for 1.5 s. The student can now drag and manipulate the molecule around the flask space. By vertically lowering the wand (using the Z axis), the student drops the molecule into the flask. This embodied gesture was designed to mimic the physical lab action of using a pipette. At its home panel, a molecule is a single sphere. When a student tosses it into the water flask it begins to move and dissociate into more spheres. These represent particles or the aqueous components of either the acid or the base. For example, when H_2SO_4 is added to the flask, it undergoes ionization by splitting into its aqueous parts, and one SO_4 (a large particle sphere and two ionized H_3O^+ [hydronium] particles).



Figure 2. Floor projection of the chemistry titration scenario.

Each molecule color matches its parent panel color, and this follows standard scientific representation of red for acids and blue for bases. Each molecule is labeled with its corresponding chemical formula. The full-surround sound is also designed to support embodied learning in a multimodal manner. Students begin to map the pitches with the chemical reactions—this differs significantly from what goes on in a regular chemistry lab. For example, a low bass tone is triggered to indicate when a new molecule is introduced to the virtual flask. If hydroxide and hydrogen particles collide to form a water molecule, a sharp ‘pinging’ tone occurs. The embodied sonic tonal feedback highlights the critical learning events: water formation, titration, and the reacted “indicator” (when the indicator returns to a stable state). With this sonic feedback, students are shaped to focus on the significant dynamic learning objectives. When the solution is fully titrated a techno-beat, that includes all the individual sounds, is played as final feedback.

To address embodiment issues for a molecular learning task, a congruent gesture was sought. Students control the speed and direction with which molecules are added, e.g., by rapidly “throwing” a molecule in using a bowling motion, the initial velocity of the molecule was greatly increased. This simulates “stirring” the solution at a faster rate, students can see how this speeds up the reaction. In contrast, a slow, downward gesture, like dipping a pipette, will gently place the molecule in the flask where it will remain relatively stationary until it collides with another. By manipulating their physical movements, students can inject kinetic energy into the system to speed up the rate at which molecules react. Visual/symbolic feedback is also included in the system. The pH updates in real time. Without molecules, the flask’s starting pH is 7.0 (neutral). As molecules in the virtual flask collide with one another, they “react” in one of four different ways based on the general chemical properties of acids and bases in an aqueous solution. The mediated environment allows students to see how each molecule affected the pH level in a step-wise manner that is not transparent during a typical wetlab since students only get final feedback when the solution turns pink.

Collaboration was designed into the system by dividing students into teams and rotating them through different roles. For example, the “acid team” and “base team” were each lead by a student who would only add those types of molecules. The teacher often stopped the action and asked the teams what their predictions would be, e.g., “What will happen to the pH if two more bases are added?” A “questioning team” was formed to lead such discussions and vote on the validity of the predictions. A handheld wireless remote was used to pause, play, and reset the scenarios. Either the teacher or lead student used this to pause support moments of classroom reflection; these may have included pauses for analysis, question-and-answer, or to retest a hypothesis. (It should be noted that a pause and reflect paradigm was also used in the regular instruction wetlab condition).

Days two and three each began with a review of concepts learned on the previous day and whole-group discussion of any questions or observations that students raised (again, similar to the control condition). If a student had a question, the teacher first turned the question over to the class looking for an explanation from peers. If students could not come up with an answer, the teacher would intervene. All students cycled through and were active in *SMALLab* by the end of each 50-min class session.

Collaboration and gameplay. On the third day, students played a competitive game with the scenario for the final 25 min of class. Either the teacher or a student would call out the goal of the game. For example, one student would populate the entire solution with bases and acids while another team was looking away. Then, the team would have to calculate how many water molecules were formed, based on the ions that were remaining. In another game, two teams of students would have to estimate how many molecules of base it would take to neutralize a given acidic solution.

Multimedia lends itself to gameplay. Gameplay is integral in our collaboration. Serious games have the explicit goal of helping students comprehend important content and may also teach “problem solving strategies, and cognitive or social skills” (Graesser, Chipman, Leeming, & Biedenbach, 2009). Such games serve a purpose beyond entertainment (Salen & Zimmerman, 2003). They do this by keeping students engaged and maintaining players’ feelings of being “pleasantly frustrated” (Gee, 2007). The goal for the main titration game was to neutralize the solution in as few “moves” as possible. A move was defined as the action of adding one molecule into the flask. This game served as a catalyst for further hypothesis development and testing as competitors challenged each other to win through efficiency and strategy. With knowledge of the chemistry system, and a preplanned strategy, it was possible to win in a few thoughtful moves. All students appeared to be deeply engaged during gameplay.

Regular instruction. Regular classroom instruction was divided between lecture and hands-on chemistry lab sessions. There was a day of preparatory lecture with slides and active discussion. The second day consisted of a hands-on titration wetlab where students worked with glassware and solutions. The teacher walked around to the small groups and engaged the students in questions and encouraged them to make predictions. Small groups of three and four worked together with the glassware, they made notes in their lab books, and discussed results within the group. On the third day, students discussed the lab experience and shared their data as a whole class. The teacher encouraged predictions and hypothesis sharing. At the end of the third day, students’ class work consisted of answering open-ended questions with several formulae included. The addition of formulae in the regular class condition is the most salient content difference between the two conditions that appeared on the tests. However, it should be noted that the students had seen the formulae for the molecules over the entire semester.

Achievement measure. The same assessment measure was used for all three test points. A sample of the test can be found in the online supplemental materials. The test included 10 major topic multiple choice questions, with eight prompts for explanations. The prompts included starter stems so students would understand that we wanted more *why* and *how* information, and not merely a rephrasing of the multiple-choice question. Lee and Songer (2003) recommended adding explanation prompts to help “push” students’ writing skills in science tests and serve as a structural support. Multiple choice answers could receive nine points each and the constructed responses to prompts were scored in the following manner: 0 (*blank, don’t know, misconception*), 1 (*partially correct*), or 2 (*totally correct*). Thus, scores could exceed 100, and only one student scored up to 114.

Informational equivalencies. For an intervention study it is crucial that the information covered in both conditions be similar and that the test not be biased to favor any one condition. Table 2 lists the informational overlap between the two conditions and the items covered on the test. Column four indicates that there were two instances of nonoverlap relating to the EMRELE condition and the test. Two test items included formulas, but no formulas were introduced during EMRELE.

Results

Tests were scored by a subject matter expert who was blind to condition. The two groups did not differ at pretest, $t(49) < 1.00$. Table 3 reports the means and effect sizes. The variances were not significantly different at pretest (Levene's test, $F < 2.00$). At Test 2 (midtest) the groups' means differed significantly; however, because the variances were also significantly different (Levene's $F = 9.28$, $p = .004$), the degrees of freedom have been adjusted. After the adjustment on the degrees of freedom, the two groups' means were still significantly different at midtest, $t(19.41) = 2.49$, $p = .02$. Effect sizes (ES) throughout this article are standardized using the unweighted average of the two groups' standard deviations.

We also analyzed midtest performance with an analysis of covariance (ANCOVA) using pretest as the covariate. The effect of intervention was significant with unadjusted variances, $F(1, 48) = 11.21$, $p = .002$. Group 1's test proportion scores were pretest = .29, *SMALLab* then midtest = .53, regular instruction then posttest = .63. For Group 2 the proportion scores were pretest = .30, regular instruction then midtest = .34, *SMALLab* then posttest = .66. The two groups' means were not significantly different at posttest, $t(49) < 1.00$.

Informational equivalencies. Table 2 reveals that the information covered in the two conditions was essentially the same. More crucially, the majority of the information covered in the test needed to be present in both conditions. Column four in Table 2 reveals the overlap between items in Regular Instruction, *SMALLab* and test item for the 10 test items. "Both" means that the concept was taught in both conditions. For this study, two items using formulas were reviewed only in the Regular Instruction condition. Thus, there may have been a small test bias favoring the regular condition, although a supplemental analysis on those two items at midtest did not reveal this to be the case ($F < 2.00$). That is, the groups were still matched at midtest on both those items. Again, this study occurred at the end of the school year, so students had been exposed to various formulas and equations throughout the year.

Discussion

The results demonstrate that whenever students were in the EMRELE condition called *SMALLab*, they learned significantly more than students in the regular instruction condition. The teacher was the same throughout the study, so teacher effects were controlled. Our partner teachers are professionals and one would expect some gains after regular instruction. Indeed, there is a small to moderate effect sizes for regular instruction. At the final learning point (Test 3) after 6 days of instruction, there was not a statistically significant difference between the two groups' perfor-

mance. This result does not support our hypothesis from the introduction section that students who received regular instruction before *SMALLab* would be at a posttest advantage.

In an earlier STEM study (Geology; Birchfield & Johnson-Glenberg, 2010), the group that received *SMALLab* as the second intervention maintained significantly higher results at test Time 3, $F(1, 64) = 3.94$, $p = .05$. This may be because mastery of that specific content relied on a more defined sequence of study. For example, knowledge regarding which fossils might reside in which depositional layers would help students operate more smoothly in the immersive space with a timer running. Perhaps in this current study—with content that was more symbolically abstract—there might be no facilitative effect for doing the wetlab experience first. In addition, after 6 days of learning about titration students may have reached individual "ceilings" for informational extraction of the content.

The order of intervention question may actually be extremely complex, and type of content may interact with order effects. We ask the question here because many teachers report feeling pressed to cover large amounts of content in a short time, but they want to do the best job possible—they want to know which type of instruction leads to greatest learning gains but also if they had more time to teach would it make sense to cover the content with both methods? This study suggests that students learn more in a short time if placed in an embodied, mixed reality learning environment called an EMRELE (midtest gain greater than 25%) compared to traditional instruction (midtest gain less than 4%). If the teacher has time for only one mode of presentation, that mode should be embodied. If there is time for more instruction, then further gains can be made using wetlab instruction. For the domain of chemistry, it does not appear that it matters which mode of instruction comes first in the sequence. The next study addresses the flexibility of the EMRELE with a new teacher and new content and attempts to replicate the previous order of intervention results.

Study 2: Disease Transmission

Method

Participants. The participants came from the same urban high school as the Study 1 students; no student had served in the previous experiment. The study began with 65 students, but only 56 were used in the analyses because seven did not complete all three tests. Two were removed from the analyses because they exhibited unusual decreases (greater than 24%) from test to test. One student was dropped from each condition. The two students had done well on the pretest; however, at midtest they both left a middle page blank. The scorers agreed that these should be considered mechanical errors and their data were not included. An equal proportion missed tests and were deleted from each class: two from Period 2, four from Period 3, and three from Period 6. All three periods were composed of 10th and 11th grade Honors biology students. The teacher had been teaching biology for 17 years.

Design. A lesson plan can be found in the online supplemental materials. The Disease Transmission study was based on the same 6-day repeated measures counterbalanced design used in Study 1. In addition, we explored student attitudes regarding technology-based embodied learning and novelty. Because this was the end of the semester, the teacher was confident that her three classes would

Table 2
Informational Equivalency Table for Study 1: Titration

Day	Scenario		Test topics	Equivalencies	
	Titration—Regular	Titration— <i>SMALLab</i>		Present in which condition	Notes on nonoverlap
1	Teacher facilitates discussion on open-ended worksheets. Difference between acids and bases. Clarifies the ionization processes via inquiry. Clarifies properties of acids. How to produce H^+ (as H_3O^+) ions in water—hydronium ion is a hydrogen ion attached to water molecule. Writes several formulas on the board.	Explore navigation in the space. Difference between acids and bases.	1. Difference between ionization and dissociation?	Both	
		Clarifies the ionization processes via inquiry.	2. Products of a neutralization reaction?	Both	
		Clarifies properties of acids. Hydronium ion is a hydrogen ion attached to water molecule. More verbal and graphic presentation.	3. At the endpoint of a neutralization reaction you will have . . . ?	Both	
	What is molarity? What is pH level?	What is pH level?	4. What is formula for the hydronium ion?	Only regular instruction	No formulas in <i>SMALLab</i>
			5. How is a hydronium ion formed?	Both	
			6. When the pH of a solution is decreasing the hydrogen ion concentration is . . . ?	Both	
2	Small groups (3 to 4) work together with glassware to titrate solutions. Students use a burette to add vinegar to discover molarity of solution. The groups discuss their hypotheses. They make predictions. The groups draw conclusions and refine conceptual model of reactions between specific ions in water, they make notes.	Introduces indicator particle. Students take turns adding the six molecules into the virtual flask and discussing how the molecules react to H_2O	7. When you add 10 ml of 1 M hydrochloric acid to 10 ml of 1M sodium hydroxide what is the net charge?	Only regular instruction	No formulas in <i>SMALLab</i>
		As a whole class they share hypotheses. They make predictions. They encourage classmate in the active space to test different molecules in the virtual flask.	8. What is the chemical process that causes an indicator to change from clear to pink?	Both	
	Use the indicator and observe how it reacts to other ions (similar to acid).	Use the indicator and observe how it reacts to other ions (similar to acid).	9. Place the five titration steps in correct order.	Both	
	Encouraged to discuss neutralization and water formation.	Discussed neutralization and water formation.	10. When a titration has reached its end point, which of the following is true . . . ?	Both	
3	Students meet as a whole class to discuss the lab experience. Each group shared their data on the titration experiment. Whole class discussed worksheets and took turns dealing with the open-ended questions.	Students meet as a whole class to discuss the lab experience. Each group shared their data on the titration experiment. Whole class discussed worksheets and took turns dealing with the open-ended questions. Spent the final half of class playing the titration games.			Same content reviewed for both groups on Day 3

Table 3
Results: Descriptives and Effect Sizes for Study 1—Titration Experiment

Group	n	Pretest		Midtest		Mid ES ^a	Posttest		Post ES ^b	Overall ES
		M	SD	M	SD		M	SD		
1 (SMALL/Regular)	16	32.83	18.19	60.62	32.65	1.09	71.44	25.17	0.37	1.78
2 (Regular/SMALL)	35	34.23	13.04	38.89	18.23	0.30	75.11	19.37	1.93	2.52

Note. ES = effect size.

^a Pretest to Midtest difference, ES SD is average of pre and mid SDs within condition. ^b Midtest to Posttest difference, ES SD is average of mid and post SDs within condition.

be matched at pretest (this was later confirmed). The classes were randomly assigned to condition. Periods 2 and 3 were grouped together (Group 1), and they received *SMALLab* first and regular instruction second. Students in Period 6 (Group 2) received regular instruction first and then *SMALLab* instruction. Unlike Study 1 (see Table S1), the pretests and posttests were administered immediately before and after the study intervention days in order to increase instructional time.

Measures.

Achievement measure. A sample of the original test is available in the online supplemental materials. This test differed from the Study 1 test in that *all* responses were constructed responses. The maximum score could exceed 100 depending on the number of diseases listed and creativity exhibited on some of the answers. Nonetheless, the highest score on the test occurred at Time 3 and was only 80.50; there was no ceiling effect.

Survey. On the final page of the posttest, a questionnaire was included to begin to examine effects of novelty. It asked the following:

How many times did you get in *SMALLab* this year?

Did it still feel new to you the last time you did it?

Have you ever done anything like it before? What was that?

Circle the number on the 1 to 5 scale that captures how much you think you learn in each situation: (a) studying the textbook, (b) classroom lecture, (c) *SMALLab* scenario.

Embodied *SMALLab* intervention.

Materials. In the Professional Learning Community meetings, the biology teacher shared several consistent misconceptions that students evidence regarding disease transmission. The new scenario focused on the concepts of: infection by viruses vs. bacteria, the difference between a vaccine and an antibiotic, the difference in time of administration between vaccines and antibiotics, and antibiotic resistance. In this scenario less emphasis was placed on the embodied aspect of design compared to the sociocollaborative aspects.

Avatars. Another promising aspect of infusing new media into education is adding user-created content to lessons. This type of participatory learning can increase engagement and the inclusion of self-created avatars has been found to be engaging for students (Falloon, 2010). Participants in the collaborative Escape from Wilson Island game reported wanting to personalize their avatars (Wendel et al., 2013). Mennecke, Triplett, Hassall, Conde, and Heer (2011) researched college students

instructed to participate in business activities, socialization, and collaboration, in the virtual environment of Second Life. His Embodied Social Presence (ESP) theory states that the body is the nexus of communication and that “an embodied representation—whether virtual, physical, imaginal, or some combination, combined with goal-directed shared activity . . . will affect the perceptions of users by drawing them into a higher level of cognitive engagement in their shared activities and communication acts” (p. 435). Participants in Mennecke et al.’s study wrote reflections on the activities and the results demonstrated that third-person inanimate pronouns describing the use of an avatar (e.g., it, its) were replaced by personal and possessive pronouns (e.g., I, me, my, his, hers) by the end of the semester. Participants also experienced a higher sense of engagement and immersion through the use of avatars. Figure 3 shows examples of some avatars created with <http://www.doppelme.com>. When students did not create avatars (approximately 25% of class by Day 1), a default monster was inserted. By the second day in *SMALLab*, all students had chosen to submit customized avatars to replace the generic monsters and so we see evidence of strong social motivations.

Figure 4 shows the interface of the floor projection with the avatars on the perimeter. The human students sit in chairs behind their avatars. The students’ primary goal is to keep their avatars alive for as long as possible by avoiding exposure to the disease while maintaining an elevated health state. The current health level and disease state for each avatar is shown in Figure 5. The outer circle depicts the health state, and this is constantly ticking down (decreasing) as time unfolds in the simulation. The inner disc color

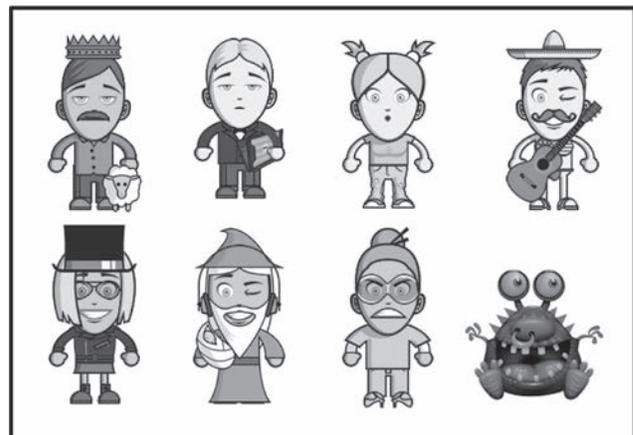


Figure 3. Examples of student-created avatars.



Figure 4. Floor projection of the disease transmission scenario.

(red, yellow, or white) indicates whether the avatar is a symptomatic carrier, asymptomatic carrier, or healthy individual, respectively. A dead avatar is depicted by skull and crossbones.

Using the tracking wand, students activate their avatar and drag it into the central space. When two avatars are co-present in the active center space there will be an invisible interaction between their disease states. This means that if neither avatar is infected, there is no change in health meter or color, but, if one avatar is a carrier, then the infection will pass to the healthy avatar in a few seconds. Despite this danger of infection, students must periodically enter the active space in order to maintain the health of their avatars (e.g., to get water) or the life force will tick away. Students drag their avatars to the active central icons (water fountain or medicine bottle) in the center of the floor. The water fountain represents sustenance. Taking the avatar to the fountain and back to its home position will refill the avatar’s health outer circle. The medicine bottle will combat infection.

The scenario lasts for 3 days. On the first day, the avatars display a slow health depletion rate—as evidenced by a slowly decrementing outer circle. The teacher speeds up the depletion rate on later days. The teacher starts the learning sequence with an exploration of how transmission occurs, students must deduce that it is when avatars are co-present in the active space. When all students have understood this, she moves on to bacterial vs. viral disease transmission and then to antibiotic resistance. On the second day in *SMALLab*, asymptomatic disease carriers are visually depicted with a yellow outer ring. However, by the third day, asymptomatic disease carriers are indistinguishable in color. This makes it more difficult for the students to deduce transmission. It is an exercise in systems level thinking to deduce which avatar is the carrier. (This maps well to an exercise in the control condition.) Students on the perimeter are highly engaged in in this scenario, they are often shouting encouragement. Because the student with the wand has the capability of

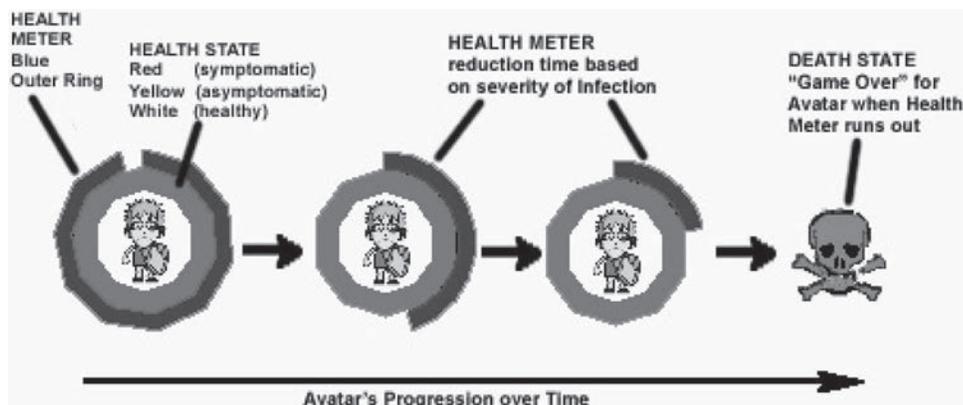


Figure 5. Example of a progression of disease affecting an avatar.

saving any avatar, students on the sides are often pleading for their avatars' lives and trying out hypotheses as to which avatar is the asymptomatic carrier.

Regular instruction. The 3 days of regular instruction consisted of 1 day of a hands-on project and 2 days of learning in a different mediated environment—viewing professional science movies on DVD. On day one, the whole class partook in a sociocollaborative project called OUTBREAK. Students needed to discover who among them was the disease carrier. This is very similar to the goal of day one in the EMRELE. The instructions were as follows:

There is a new epidemic spreading throughout the continent at an incredible rate. This outbreak has been caused by the pathogen *Stumpfacyllus stephancisus*.^a It is spread from person to person through casual contact. Having a conversation with just one infected host can cause mental meltdown, twitchy fingers, an insatiable thirst for "Sunny Delight," uncontrolled fully body hair growth, and tooth loss. Be careful!

^aThis is a play on the teacher's name.

Each student was issued a test-tube of clear liquid. One of the test-tubes contained a small amount of bleach. Students needed to approach up to four other students and pour the liquid into each other's vials. They kept notes on contact. The teacher created four groups—she walked around and inserted a drop of a solution into each test-tube, the ones that contained some bleach turned pink (signifying an infected state). The four teams needed to collaborate to deduce who was the original carrier. Students were very engaged during this project. On Day 2, the teacher gave a summary lecture on the topic of symptomatic and asymptomatic carriers. Students then watched a professionally produced DVD from National Geographic's "The Virus Hunters" (Elisco & Biega, 2009). This Explorer series documentary investigated a new theory suggesting that all life—including humans—descended from viruses. The DVD presented content relating to viruses and transmission vectors. On day three, students watched a PBS show called "Evolution: Evolutionary Arms Race" (Apsell & Ritsko, 2001). It included a discussion of supergerms, deadly microorganisms, and antibiotic resistance. The tuberculosis epidemic was used as an example of how an infectious disease can become resistant to treatment. The students were required to take notes during both films, and key elements were discussed at the end of each class.

Results

Two scorers, blind to condition, scored the three tests. Interrater reliability on a random sample of 74 items revealed a significant

correlation ($r = .94, p < .001$). The two groups did not differ at pretest ($F < 1.0$). Table 4 lists the descriptive statistics with ES's for the three test points.

An ANCOVA with pretest as the covariate revealed that the groups' midpoint tests were marginally different, $F(1, 53) = 3.07, p = .084$, with the *SMALLab* group being favored. To assess whether the gains at test Time 3 differed by condition we needed to adjust for the new start point (i.e., midtest) and so a t test was analyzed using the gain scores from midtest (Time 2) to posttest (Time 3). This time the result significantly favored Group 2, the group that received *SMALLab*, $t(53) = 2.39, p = .021$. The final test points were not significantly different. The proportions correct on the tests were as follows: Group 1 pretest = .39, *SMALLab* then midtest = .47, regular then posttest = .49; Group 2 Pretest = .36, regular then midtest = .40, *SMALLab* then posttest = .48.

Informational equivalencies. An informational equivalency analysis was also made between the content taught in the two conditions and the test items. Column four in Table 5 reveals the overlap between items in Regular Instruction, *SMALLab*, and test item for the 12 test items. "Both" means that the concept was taught in both conditions. For this study, one item was covered only in regular instruction, one item was covered only in *SMALLab*, and one item was covered in neither condition; overall the test content was not biased to favor either condition.

Survey/questionnaire. We wanted to begin to assess how much of the learning may have been due to novelty.

1. *How many times were you in SMALLab this year?* The frequencies in the space ranged from 1 to 10, the majority were in 2 to 4 times.
2. *Have you ever done anything like this before?* Only one student said "yes, Wii."
3. *How much do you think you learn in each situation?* The students circled a number ranging from 1 to 5, with 5 being the most. The means and *SD* were: *SMALLab* 3.85 (0.89); Lecture 3.32 (0.96); Textbook 2.79 (1.06). Using a Wilcoxon signed ranks test for related samples, 28 out of 53 students reported they learned more in *SMALLab* compared to attending a lecture ($Z = 2.69, p = .007$). For the comparison between *SMALLab* and reading a textbook, 36 out of 53 reported they learned more in *SMALLab* ($Z = 4.24, p < .000$).
4. *Did it still feel new to you the last time you did it?* This question was intended to quantify novelty for the participant. There was an even split on this question. Of the 51

Table 4
Results: Descriptives and Effect Sizes for Study 2—Disease Transmission

Group	<i>n</i>	Pretest		Midtest		Mid ES ^a	Posttest		Post ES ^b	Overall ES
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>		
1 (SMALL/Regular)	39	38.99	16.48	47.19	14.46	0.53	48.58	15.23	0.09	0.61
2 (Regular/SMALL)	17	35.85	13.44	40.17	13.37	0.32	48.33	11.30	0.66	1.01

Note. ES = effect size.

^a Pretest to Midtest difference, ES *SD* is average of pre and mid *SD*s within condition. ^b Midtest to Posttest difference, ES *SD* is average of mid and post *SD*s within condition.

Table 5
Informational Equivalencies for Study 2—Disease Outbreak

Day	Disease—Regular	Disease— <i>SMALLab</i>	Test topics	Equivalencies
1	Perform hands-on disease project with test tubes and bleach Generate hypotheses about who was the carrier. Deduce carrier. Do they always show symptoms? Discuss symptomatic and asymptomatic carriers. What is difference between bacterial and viral? What is a quarantine?	Explore the space and navigation. Ascertain how the virtual disease is transmitted—through the air, the water? Wand in same area? Work to keep yourself alive by not being in space with others who are sick—similar to a quarantine. Discuss symptomatic and asymptomatic carriers. What does it mean to have a yellow ring and be a carrier?	1. What is a quarantine? 2. What are bacteria? List some diseases. 3. What is a virus? List some diseases. 4. What is a vaccine? 5. What are antibiotics? 6. A number of microbes in a dish will grow . . . 7. What is the difference between being infected and being symptomatic 8. What are the major differences between vaccines and antibiotics? 9. What is antibiotic resistance?	Both Both Both Both Regular instruction only
2	Review disease transmission project from yesterday. Students watched “The Virus Hunters” video. Investigated a new theory suggesting that all life is descended from viruses. Went into detail on virus definition and mechanisms. Discussed transmission vectors and how the spread can occur. Difference between bacterial and viral infections. Student took notes then discussed movie at end of class. Focused on difference between vaccines versus antibiotics.	Further explore difference between symptomatic and nonsymptomatic carriers. Increase number of students who are in space at the same time and rate of infection. Students must work with more social cohesion, figure out whom to exclude from common area. Limit medicine supply—discuss how this changes system dynamics. Discuss difference between viral and bacterial infections. Introduce antibiotic resistance. Vary antibiotic resistance threshold	10. There is an outbreak of bacterial plague in your village—what would you do? 11. Logic question re: waiting for virus to die before accessing food and water again 12. Make a linear graph—On Day 1 four people are sick. On Day 10 forty people are sick. How many are sick are on Day 5?	Both Both Both Mainly practiced in <i>SMALLab</i> Neither condition
3	Students watched “Evolutionary Arms Race,” took notes and discussed at the end of class. Discussed supergerms, deadly microorganisms, and antibiotic resistance. The tuberculosis epidemic was analyzed in depth for disease resistance. Mentioned that disease-causing microorganisms are modern humans’ only predators. Antibiotic resistance and adaptation are covered. Students took notes and discussed key topics at the end of class.	Review and reinforce antibiotic resistance and limited medicine supplies. Explore being judicious about timing of medicines. Formally introduce vaccines. Alternate between vaccination and antibiotic resistance simulations. End with game rounds that limited medicine supply and water/health supply.		

who answered this question, 26 reported it did *not* feel new at the last time, and 25 reported it still felt new. This response was not significantly predicted by number of times in the space, using yes/no new as the DV, and number of times in the space as the predictor in a logistic regression (Wald = 2.02, $p = .16$).

Discussion

The results of the disease transmission study reveal the same pattern of results seen in Study 1: (a) Groups begin with equivalent knowledge; (b) at midtest the *SMALLab* first group learns more than regular instruction first; (c) at posttest the two groups converge after both receive instruction in *SMALLab*; and (d) each time a group received the EMRELE (*SMALLab*) intervention, larger gains in learning were seen. Interestingly, this study resulted in smaller effect sizes (gains) compared to the previous titration study. There may be several reasons for this.

First, this scenario is very ambitious at the scale level. Students must conceptualize at both the micro-level of unseen disease organisms *and* the macro-level of contagion effects on a society. Second, the scenario requires rapid decision making about multiple abstract, nonvisible factors concerning transmission vectors, medicine types, and quarantine costs and benefits. Complex systems level thinking skills are required while attending to powerful social cues—does one save only one's friends? By contrast, students in Study 1 (titration) were able to stop the action and assess what was happening as large graphic molecules collided and reacted with one another—a localized effect that was not under time pressure. However, we suspect the primary reason for the differing effect sizes is that the disease transmission instrument was structured differently than the titration instrument. The disease test was composed almost entirely of constructed responses. This was a more difficult test than the titration test and assessed a more open-ended domain, thus score range was constricted. The maximum score on the titration test was 114.00; the maximum score on the disease test was 80.50. Even though the gains were smaller in Study 2, they are in the same direction (and each study results in a slanted diamond-shaped graphic with the two group lines diverging at midpoint and reconverging at posttest), and they support the premise that mixed reality, immersive learning environments that afford strong group-level activities can result in larger learning gains compared to regular instruction. This holds even though regular instruction in Experiment 2 included a day of interactive project-based, small group collaborative learning.

General Discussion

(a) *Do students learn more in an embodied mixed reality learning environment compared to regular instruction?* The data from both studies provided a clear, affirmative answer. Each time students were in the computer-based, embodied condition they demonstrated moderate to large gains in learning (effect sizes of 0.53 to 1.93); each time students were in the regular instruction condition they demonstrated small to moderate gains in learning (effect sizes of 0.09 to 0.37).

The answer to the second question—(b) *If students were to receive both types of instruction, does order of intervention matter?*—may not yet have a conclusive answer. The data from these

two studies are consistent with each other in that the order of intervention did not affect final outcome scores; however, these results are not consistent with the earlier cited Birchfield and Johnson-Glenberg (2010) study. In that Geology study, the students in the regular instruction before the EMRELE condition demonstrated significantly greater learning gains at the final posttest. It may be that certain content needs a solid foundation before being introduced in a more sociocollaborative, technology-mediated platform. A deeper content analysis coupled with item response analyses might illuminate the conditions under which regular instruction should precede EMRELE mixed reality lessons. Gire et al. (2010) found complicated test item type by order of intervention effects. They investigated learning differences using real and virtual pulleys in a waitlist design and found that students learned different content better depending on order of intervention. They attributed these differences to the physics concepts' salience and the match with condition. That is, the real, tangible pulleys gave students a "kinesthetic experience" with effort and distance pulled; this was not so salient for the more abstracted physics concept of work, and the virtual pulleys were not designed to be gesturally congruent.

Reasons for the Gains in Learning

Embodiment. In these two studies, three out of four times when the technology-based EMRELE platform was compared to regular instruction, significant gains were seen favoring the embodied EMRELE learning environment; in the fourth instance it was a statistical trend. We posit that the level of embodiment in the lessons is of primary importance in explaining these gains. Chemistry is typically an abstract, formula-rich domain, and many students struggle with it. As Barsalou (2008) states, little empirical evidence supports the presence of amodal symbols in cognition. Amodal symbols provide powerful formalisms for representing knowledge and simulating artificial intelligence, but for human learners such symbols may not provide the most felicitous path to comprehension. If the cognitive primitives (in the sense of diSessa's, 1983, scientific P-prims) are primarily embodied, then learning environments that highlight physically grounded knowledge and that induce mental simulations of constructs like chemical reactions and disease avoidance tactics, may be more powerful learning methods than techniques emphasizing observation and symbol memorization. Bodily perception and action, and the experiences based on perception and action provide a mechanism for grounding. The two EMRELE scenarios described in this article did not use mathematical equations to teach the larger concepts, instead they enlisted several modalities for learning (visual, auditory, and kinesthetic). Our position is that the more modalities and well-mapped, congruent afferent sensorimotor activations that are recruited during the encoding of the information, then the crisper and more stable the knowledge representations should be in schematic storage. Our latest research also demonstrates that science content (centripetal force) learned in a high embodied condition compared to a low embodied, observational condition is remembered better during 1-week follow-up tests (Johnson-Glenberg et al., 2013).

Although regular instruction in titration included a hands-on lab where students used physical tangibles to mix solutions, the chemical processes were never rendered explicitly. In contrast, while

learning in the embodied titration scenario, students were making choices about where to place the molecules and with which velocity. They immediately saw how the reactions occurred, they heard ongoing auditory feedback regarding interactions, they had the ability to stop the dissociation process and reflect. In addition, they had the ability to walk around and view the titration process from different angles. We cannot point to which exact kinesthetic, or multi-modal act proved to be most predictive of the learning gains, nor can we disentangle the effects of the co-located group collaboration because these EMRELE lessons are complex, real classroom learning situations. The study is important because it demonstrates, in two different science domains, that when using all the affordances of an immersive, digitized EMRELE platform educators can expect to see learning gains above and beyond quality traditional pedagogy that includes hands-on and group-based components.

From the designer's perspective, creating a module with the highest or fourth degree of embodiment means that the module should encourage the student to physically activate a large quantity of sensorimotor neurons in a manner that is congruent to the content being learned. We should be striving to create fourth and third degree of embodiment lessons that encourage students to get out of their seats and be generative. The focus of our discussion of embodied learning thus far pertains to students who are physically moving in the space, what role does *observing* someone else's embodied learning play.

Recent fMRI and single cell experiments suggest that neurocorrelates of the brain are active during both observational and active procedures (Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010; Pulvermüller, 2005). However, a relative attenuation of learning may occur during observational learning because observing does not engage the same magnitude of activation in the neuromuscular system. Observational learning can still be considered embodied, but low embodied, first or second degree, as long as the content is not simply symbol manipulation. That is, content should contain perspectives and graphics that might trigger our mirror neuron systems (Rizzolatti & Craighero, 2004) for observational embodied learning. Engelkamp (2001) verifies that the self-performed task effect size is always greater than the experimenter-performed task effect size (i.e., the observational task). We hold that by *doing the action*, by engaging more sensorimotor activation in a gesturally congruent manner a student will be learning to a higher degree of embodiment.

Collaboration. We posit that collaboration is another important causal factor associated with learning gains seen in this EMRELE. Science labs can be highly collaborative with activities designed to advance learning by harnessing students' innate social and collaborative capabilities (Roth, Woszczyzna, & Smith, 1996). Using the example of the disease transmission study, we created a highly collaborative, co-located experience; students experienced the disease transmission process face-to-face and their physical proximity to one another played a critical role. Decision-making strategies regarding whom to save, the ramifications of medicine distribution, and the cost-benefit contingencies of staying out of the central area even when water was needed (a type of quarantine) were situated in synchronous game play. We believe some of the learning gains may be attributed to two of the standard collaborative learning perspectives, the first is social cohesion (success of the group as a whole is important, in addition to the performative

motivation in our scenarios), and the second is the cognitive developmentalist perspective (where the more advanced students would talk aloud, answer teacher questions, and model behaviors that others could learn from). It should be noted that in both studies the regular classroom instruction used collaborative small groups to run hands-on experiments, but there was not an emphasis on whole classroom collaboration. Another bonus of adding virtual and mixed reality components to a lesson may be that multiple trials of disease transmission could occur, up to eight a day, while in the bleach-and-test tube regular instruction lesson only one trial could occur. The EMRELE afforded more opportunities for fresh discourse and collaboration.

It is not always a given that collaboration will result in higher learning gains. Nihalani et al. (2011, Experiment 1) taught undergrads how to navigate through an educational computer simulation (that would be considered second degree in our proposed taxonomy). Students who received individual feedback outperformed students placed in three-person collaborative triads with feedback. A further complicating factor is that students arrive to the content with varying level of expertise. In addition, most people would predict that human tutor dialogue would always result in greater learning gains, compared to canned or text-based dialogue, but a physics-based study (van Lehn et al., 2007), demonstrated that only the low prior knowledge students working on content that was above their skill levels were differentially aided by the human tutoring help. There are also subtle intragroup social effects that can affect learning. Wendel et al.'s (2013) serious game study found that teams with one person acting as a leader during collaborative tasks (e.g., three players must carry a virtual palm tree) performed significantly better than teams without a leader. More research is needed on the effects of collaboration in EMRELE platforms.

Mediators of language and novelty. Due to space constraints we did not report on evidence that students' language is altered by the embodied EMRELE environment. By allowing students to participate in a space designed for guided discovery and the free sharing of ideas, more opportunities emerge for peer-to-peer collaboration in the science classroom. In this titration study, we found that student language stayed on task (100%) during the EMRELE condition, while in the small group setting one third of language turns were devoted to off-task content.

Finally, we cannot ignore the novelty of the technology. Humans are designed to be attentive and responsive to novel stimuli. Novelty induces a very rapid orienting response (Lisman & Grace, 2005). Novelty motivates us to attend to the incoming information because such information holds the promise of a reward. Novelty also activates the limbic system as well as attentional systems. Tulving, Markowitsch, Craik, Habib, and Houle (1996) postulated that the novelty of the information predicts whether it will make it into long term memory (i.e., knowledge schema). Given the neural predisposition to attend to novelty, perhaps it should not be treated as a nuisance variable in education—perhaps we should design with it in mind, try to maximize it for each lesson to keep students engaged? At issue for a recurring learning environment is whether the novelty will fade. In Study 2, we included a post survey question that asked whether *SMALLab* still felt "new" the last time a student had entered it. There was an even split in the responses. However, the number of times students were in the space was not predictive of the response. The experience of novelty may be a

subject variable—and perhaps a potent mediator—that needs to be further explored. We believe that even if the overall environment of *SMALLab* itself fades in novelty, designers should strive to keep each embodied lesson as fresh and as gesturally congruent as possible.

Final note on games and performative aspects. Serious games are beginning to filter into the classroom, we think this is good news and hope curriculum designers create games that are as embodied as possible. Although a meta-analysis by Young et al. (2012) revealed learning effects for only exergames, history and language learning games, the larger and more recent Wouters et al. (2013) meta-analysis found strong effects in all domains except biology and engineering. Wouters et al. also predicted and found more durable learning associated with serious games compared to conventional instructional methods (delayed test Cohen's $d = 0.36$). All of our current studies now include delayed tests.

Finally, the performative aspects of a large scale EMRELE should be noted. Students observing on the sidelines know that very shortly they will be called upon to perform in front of their peers. The desire to do well in front of the group may further drive engagement and hence learning. Students are watching and probably mentally rehearsing what they will do when it is their turn. We also note that for some students getting up in public and making decisions (and sometimes errors) in the EMRELE resulted in hesitancy, and perhaps some level of embarrassment.

Future Directions

Movement as learning predictor. Because it is now possible to capture and record in a cost-effective manner movement in 3D space, researchers should be gathering metrics on movement while learning. We are now creating tests and learning exercises that incorporate gesture as an assessment metric gathered with Microsoft *Kinect* and will soon be exploring the LEAP Motion technology as well. There is a rich and growing literature on gesture in education (Alibali & Nathan, 2011; Cook et al., 2010). Schwartz and Black (1996) argued that spontaneous hand gestures are “physically instantiated mental models.” In a study on interlocking gears, they found that participants gestured the movement of the gears with their hands to help them imagine the correct direction of the gears and in this way the participants gradually abstracted the turning rules. We view gesture as strong sensorimotor grounding when it is congruent with the content. In addition, when students are allowed to use their own metaphoric or iconic gestures these may serve as both primes and memory retrieval cues. More research is needed to ascertain whether designing educational content with gestures that promote the correct, congruent kinesthetics is a more effective method for learning compared to more traditional methods.

Unfortunately, these two studies did not have enough statistical power to run aptitude by treatment interactions (ATI) analyses. One of our predictions for future EMRELE studies would be that low prior knowledge (PK) learners benefit more from having fewer choices in the space. For example, there is complex systems-level thinking involved in figuring out who is an asymptomatic carrier in a space filled with 20 moving avatars. A low PK student might need to work his or her way up to that level of complexity much slower than a higher PK student. The user interface may

need to be altered, such that low PK students only work with four active avatars in the beginning and less sonic feedback.

Assessments. Besides adding in-process gross movement and gesture as metrics, we also need to be more creative about pretest and posttest measures. We should move beyond pure text-based responses and gather motion metrics during learning and allow participants to draw images freehand on tests. Lindgren and Moshell (2011) found that participants in their embodied meteor platform installation were more likely to include dynamic elements (arrows, etc.) in their astronomy drawings compared to participants who learned via desktop computer version.

Conclusions

Technology-based learning is typically implemented using desktop interfaces and increasingly tablet-sized screens. These configurations are designed around a single-user model, and so the computerized science lab can, in some cases, lead to an isolating experience that runs counter to students' highly sociocollaborative experiences with other physical tools. Immersive, mixed reality learning environments can be “technological” and collaborative. These EMRELE environments can foster and support quality face-to-face, co-present collaboration with digitized components. These two studies support the hypothesis that technology-based, embodied media can be both effective and collaborative, and these types of environments can support established classroom curricula.

In addition, this article proposes a taxonomy for embodied learning in educational spaces. Fourth degree, or the highest level of embodied learning is driven by three components: the amount of sensorimotor activation, gestural congruency with content, and perception of immersion. When properly designed, the use of EMRELEs can result in significant learning gains.

References

- Alibali, M. W., & Nathan, M. J. (2011). Embodiment in mathematics teaching and learning: Evidence from students' and teachers' gestures. *Journal of the Learning Sciences*. Advance online publication. doi: 10.1080/10508406.2011.611446
- Apsell, P. S. (Producer), & Ritsko, A. (Managing Director). (2001). *Evolution: The evolutionary arms race* [Video]. Available at http://www.pbs.org/wgbh/nova/education/programs/0000_evoevoar.html
- Asllani, A., Ettkin, L., & Somasundar, A. (2008). Sharing knowledge with conversational technologies: Web logs versus discussion boards. *International Journal of Information Technology and Management*, 7, 217–230. doi:10.1504/IJITM.2008.016607
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639
- Birchfield, D., & Johnson-Glenberg, M. C. (2010). A next gen interface for embodied learning: *SMALLab* and the geological layer cake. *International Journal of Gaming and Computer-Mediated Simulation*, 2, 1, 49–58. doi:10.4018/jgcms.2010010105
- Birchfield, D., & Megowan-Romanowicz, C. (2009). Earth science learning in *SMALLab*: A design experiment for mixed-reality. *Journal of Computer Supported Collaborative Learning*, 4, 403–421. doi:10.1007/s11412-009-9074-8
- Black, J. B., Segal, A., Vitale, J., & Fadjo, C. (2012). Embodied cognition and learning environment design. In D. Jonassen & S. Lamb (Eds.), *Theoretical foundations of student-centered learning environments* (2nd ed., pp. 198–223). New York, NY: Routledge.

- Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., & Witherington, D. (2000). Travel broadens the mind. *Infancy, 1*, 149–219. doi:10.1207/S15327078IN0102_1
- Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*, 73–105. doi:10.1111/j.1756-8765.2008.01005.x
- Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of Memory and Language, 63*, 465–475. doi:10.1016/j.jml.2010.07.002
- Crook, C. (1996). *Computers and the collaborative experience of learning*. London, England: Psychology Press.
- Cruz-Niera, C., Sandin, D. J., & Defanti, T. (1993). Surround sound screen projection-based virtual reality: The design and implementation of the CAVE. *Computers & Graphics, 13*, 135–142.
- diSessa, A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 15–33). Hillsdale, NJ: Erlbaum.
- Dynarski, M., Agodini, R., Heavyside, S., Novak, T., Carey, N., Campuzano, L., . . . Sussex, W. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort* (NCEE Publication No. 20074005). Retrieved from <http://ies.ed.gov/ncee/pubs/20074005/>
- Elisco, D. (Writer), & Biega, C. (Director). (2009). The virus hunters [DVD]. In J. Halperin (Executive producer), *National geographic explorer*.
- Engelkamp, J. (2001). Action memory: A system oriented approach. In H. D. Zimmer, R. L. Cohen, & M. J. Guynn (Eds.), *Memory for action: A distinct form of episodic memory?* (pp. 49–96). Oxford, England: Oxford University Press.
- Engelkamp, J., & Zimmer, H. D. (1985). Motor programs and their relation to semantic memory. *German Journal of Psychology, 9*, 239–254.
- Falloon, G. (2010). Using avatars and virtual environments in learning: What do they have to offer? *British Journal of Educational Technology, 41*, 108–122. doi:10.1111/j.1467-8535.2009.00991.x
- Gallese, V., Gernsbacher, M. A., Heyes, C., Hickok, G., & Iacoboni, M. (2011). Mirror neuron forum. *Perspectives on Psychological Science, 6*, 369–407. doi:10.1177/1745691611413392
- Gee, J. P. (2007). *Good video games + good learning: Collected essays on video games, learning, and literacy*. New York, NY: Peter Lang.
- Gibson, J. J. (1979). *The ecological approach to vision perception*. Boston, MA: Houghton Mifflin.
- Gire, E., Carmichael, A., Chini, J., Rouinfar, A., Rebello, S., & Puntambekar, S. (2010). The effects of physical and virtual manipulatives on students' conceptual learning about pulleys. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Learning in the disciplines: Proceedings of the 9th International Conference of the Learning Sciences* (pp. 937–943). Chicago, IL: ICLS.
- Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 586–596. doi:10.1002/wcs.55
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558–565. doi:10.3758/BF03196313
- Glenberg, A. M., Sato, M., & Cattaneo, L. (2008). Use-induced motor plasticity affects the processing of abstract and concrete language. *Current Biology, 18*, R290–R291. doi:10.1016/j.cub.2008.02.036
- Goldin-Meadow, S. (2011). Learning through gesture. *WIREs Cognitive Science, 2*, 595–607. doi:10.1002/wcs.132
- Graesser, A. C., Chipman, P., Leeming, F., & Biedenbach, S. (2009). Deep learning and emotion in serious games. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 81–100). New York, NY: Routledge, Taylor & Francis.
- Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point& Query, AutoTutor, and iSTART. *Educational Psychologist, 40*, 225–234. doi:10.1207/s15326985ep4004_4
- Hatton, S., Birchfield, D., & Megowan, M. C. (2008, August). *Learning metaphor through mixed-reality game design and game play*. Paper presented at the ACM sandbox conference, Los Angeles, CA.
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15*, 495–514. doi:10.3758/PBR.15.3.495
- Johnson, D. W., & Johnson, H. (1991). *Learning together and alone: Cooperation, competition, and individualization*. Englewood Cliffs, NJ: Prentice Hall.
- Johnson, D. W., & Johnson, R. T. (1984). *Cooperative learning*. New Brighton, MN: Interaction Book Company.
- Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Edina, MN: Interaction Book Company.
- Johnson, R. T., & Johnson, D. W. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*, 365–379.
- Johnson-Glenberg, M. C. (2005). Web-based training of metacognitive strategies for text comprehension: Focus on poor comprehenders. *Reading and Writing: An Interdisciplinary Journal, 18*, 755–786.
- Johnson-Glenberg, M. C. (2007). Web-based reading comprehension instruction: Three studies of 3D-readers. In D. McNamara (Ed.), *Reading comprehension strategies: Theory, interventions, and technologies* (pp. 225–241). Mahwah, NJ: Erlbaum.
- Johnson-Glenberg, M. C. (2012). *What is learning in a mixed reality environment and what does an "embodied lesson" mean*. November, 2012. ARIELSig, Special Group of AERA. Supernews Website: <http://www.arielsig.net/what-is-learning-in-a-mixed-reality-environment-and-what-does-an-embodied-lesson-mean/>
- Johnson-Glenberg, M. C., Birchfield, D., Megowan, C., Tolentino, L., & Martinez, C. (2009). Embodied games, next gen interfaces, and assessment of high school physics. *International Journal of Learning and Media, 1* (). Retrieved from <http://ijlm.net/>
- Johnson-Glenberg, M. C., Birchfield, D., & Usyal, S. (2009). SMALLab: Virtual geology studies using embodied learning with motion, sound, and graphics. *Educational Media International, 46*, 267–280. doi:10.1080/09523980903387555
- Johnson-Glenberg, M. C., Megowan-Romanowicz, C., Glenberg, A., Birchfield, D., & Savio-Ramos, C. (2013). *Learning platform and amount of embodiment designed into a lesson affect physics comprehension*. Manuscript submitted for publication.
- Kaku, M. (2011). *Physics of the future: How science will shape human destiny and our daily lives by the year 2100*. New York, NY: Doubleday.
- Klahr, D., Triona, L. M., & William, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching, 44*, 183–203.
- Lakoff, G., & Nunez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York, NY: Basic Books.
- Lee, H. S., & Songer, N. B. (2003). Making authentic science accessible to students. *International Journal of Science Education, 25*, 1, 1–26.
- Lindgren, R., & Moshell, J. M. (2011). Supporting children's learning with body-based metaphors in a mixed reality environment. In P. Blikstein & P. Marshall (Eds.), *Proceedings of the 10th annual conference on interaction design and children* (pp. 177–180). Ann Arbor, MI: IDC.
- Lisman, J. E., & Grace, K. (2005). The hippocampal-VTA loop: Controlling the entry of information into long-term memory. *Neuron, 46*, 703–713. doi:10.1016/j.neuron.2005.05.002
- Liu, D., Jenkins, S. A., Sanderson, P. M., Fabian, P., & Russell, W. J. (2010). Monitoring with head-mounted displays in general anesthesia: A clinical evaluation in the operating room. *Anesthesia and Analgesia, 110*, 1032–1038. doi:10.1213/ANE.0b013e3181d3e647

- Llewellyn, D. (2005). Constructing an understanding of scientific inquiry. In *Teaching high school science through inquiry: A case study approach* (pp. 205–218). Rochester, NY: Corwin Press.
- Mennecke, B. E., Triplett, J. L., Hassall, L. M., Conde, Z. J., & Heer, R. (2011). An examination of a theory of embodied social presence in virtual worlds. *Decision Sciences, 42*, 413–450. doi:10.1111/j.1540-5915.2011.00317.x
- Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE, Transactions on Information Systems, E77-D* ().
- Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology, 20*, 750–756. doi:10.1016/j.cub.2010.02.045
- Newman, J. (2008). *Developmental perspectives on embodiment and consciousness: The Jean Piaget symposium series*. Boca Raton, FL: CRC Press.
- Nihalani, P. K., Mayrath, M., & Robinson, D. H. (2011). When feedback harms and collaboration helps in computer simulation environments: An expertise reversal effect. *Journal of Educational Psychology, 103*, 776–785. doi:10.1037/a0025276
- Noice, H., & Noice, T. (2006). What studies of actors and acting can tell us about memory and cognitive functioning. *Current Directions in Psychological Science, 15*, 14–18. doi:10.1111/j.0963-7214.2006.00398.x
- Pausch, R., Proffitt, D., & Williams, G. (1997). Quantifying immersion in virtual reality. In *Proceedings of the 24th annual conference on computer graphics and interactive techniques*. New York, NY: ACM Press.
- Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.). New York, NY: International Universities Press. (Original work published 1936) doi:10.1037/11494-000
- Plass, J. L., Homer, B. D., & Hayward, E. (2009). Factors for educationally effective animations and simulations. *Journal of Computing in Higher Education, 21*, 31–61. doi:10.1007/s12528-009-9011-x
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience, 6*, 7, 576–582. doi:10.1038/nrn1706
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience, 11*, 5, 351–360. doi:10.1038/nrn2811
- Rizzolatti, G., & Craighero, L. (2004). The mirror -neuron system. *Annual Review of Neuroscience, 27*, 169–192. doi:10.1146/annurev.neuro.27.070203.144230
- Roth, M., Woszczyzna, C., & Smith, W. (1996). Affordances and constraints of computers in science education. *Journal of Research on Science Teaching, 33*, 995–1017.
- Rouet, J., Schnotz, W., & Lowe, R. (2008). Understanding multimedia documents: An introduction. In J.-F. Rouet, R. Lowe, & W. Schnotz (Eds.), *Understanding multimedia documents* (pp. 1–14). Berlin, Germany: Springer. doi:10.1007/978-0-387-73337-1_1
- Salen, K., & Zimmerman, E. (2003). *Rules of play*. Cambridge, MA: MIT Press.
- Savvides, P., Tolentino, L., Johnson-Glenberg, M. C., & Birchfield, D. (2010, April). *A mixed-reality game to support communication for students with autism*. Poster presented at SITE: Society for Information Technology & Teacher Education, San Diego, CA.
- Schnotz, W., & Kuerschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*, 469–508. doi:10.1007/s10648-007-9053-4
- Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science, 20*, 457–497.
- Segal, A., Black, J., & Tversky, B. (2010, November). *Do gestural interfaces promote thinking? Congruent gestures promote performance in math*. Paper presented at the 51st annual meeting of the Psychonomic Society, St. Louis, MO.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153–189. doi:10.3102/0034654307313795
- Smaldino, S. E., Lowther, D. L., & Russell, J. D. (2008). *Instructional technology and media for learning* (9th ed.). Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Texas Center for Educational Research. (2009). *Evaluation of the Texas technology immersion pilot: Final outcomes for a four-year study*. Retrieved from http://www.tcer.org/research/etxtip/documents/y4_etxtip_final.pdf
- Tolentino, L., Birchfield, D., Megowan-Romanowicz, C., Johnson-Glenberg, M. C., Kelliher, A., & Martinez, C. (2009). Teaching and learning in the mixed-reality science classroom. *Journal of Science Education and Technology, 18*, 501–517. doi:10.1007/s10956-009-9166-2
- Tulving, E., Markowitsch, H. J., Craik, F. I. M., Habib, R., & Houle, S. (1996). Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cerebral Cortex, 6*, 71–79. doi:10.1093/cercor/6.1.71
- van Lehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*, 4, 197–221. doi:10.1080/00461520.2011.611369
- van Lehn, K., Graesser, A., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). Natural language tutoring: A comparison of human tutors, computer tutors, and text. *Cognitive Science, 31*, 3–52.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Methodological Issues in Researching CSCL Computers & Education, 46*, 71–95.
- Wendel, V., Gutjahr, M., Goebel, S., & Steinmetz, R. (2013). Designing collaborative multiplayer serious games. *Education and Information Technologies, 18*, 287–308.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625–636.
- Winters, A. (2008). Emotion, embodiment, and mirror neurons in dance/movement therapy: A connection across disciplines. *American Journal of Dance Therapy, 30*, 84–105.
- Wouters, P., van Nimwegen, C., Oostendorp, P., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*, 249–265.
- Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., . . . Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research, 82*, 61–89.

Received December 26, 2010

Revision received May 29, 2013

Accepted June 24, 2013 ■