Computers & Education 106 (2017) 150–165

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment

Susan Jang [a], Jonathan M. Vitale [b, *], Robert W. Jyung [c], John B. Black [d]

[a] 436 Grace Dodge Hall, Department of Human Development, Teachers College, Columbia University, New York, NY 10027 United States
[b] 4407 Tolman Hall, Graduate School of Education University of California Berkeley, California 94720, United States
[c] 90 Bergen Street, Department of Otolaryngology, Rutgers New Jersey Medical School, Newark, NJ, 07101, United States
[d] 551b Grace Dodge Hall, Department of Human Development, Teachers College, Columbia University, New York, NY 10027, United States

ARTICLE INFO

Article history:
Received 14 July 2016
Received in revised form 1 December 2016
Accepted 15 December 2016
Available online 19 December 2016

Keywords:
Virtual reality
Embodied cognition
Medical education
Direct manipulation
Spatial ability

ABSTRACT

With the advancement of virtual reality (VR) technologies, medical students may now study complex anatomical structures in three-dimensional (3-D) virtual environments, without relying solely upon high cost, unsustainable cadavers or animal models. When coupled with a haptic input device, these systems support direct manipulation and exploration of the anatomical structures. Yet, prior studies provide inconclusive support for direct manipulation beyond passive viewing in virtual environments. In some cases, exposure to an “optimal view” appears to be the main source of learning gains, regardless of participants’ control of the system. In other cases, direct manipulation provides benefits beyond passive viewing. To address this issue, we compared medical students who either directly manipulated a virtual anatomical structure (inner ear) or passively viewed an interaction in a stereoscopic, 3-D environment. To ensure equal exposure to optimal views we utilized a yoked-pair design, such that for each participant who manipulated the structure a single matched participant viewed a recording of this interaction. Results indicate that participants in the manipulation group were more likely to successful generate (i.e., draw) the observed structures at posttest than the viewing group. Moreover, manipulation benefited students with low spatial ability more than students with high spatial ability. These results suggest that direct manipulation of the virtual environment facilitated embodiment of the anatomical structure and helped participants maintain a clear frame of reference while interacting, which particularly supported participants with low spatial ability.

1. Introduction

Virtual reality (VR) systems allow users to explore immersive, three-dimensional (3-D) environments from any location, which could have a profound impact on science education (Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014).

---

* Corresponding author.
E-mail addresses: jangsusan@gmail.com (S. Jang), jonvitale@berkeley.edu (J.M. Vitale), jyungrw@njms.rutgers.edu (R.W. Jyung), black@tc.columbia.edu (J.B. Black).

http://dx.doi.org/10.1016/j.compedu.2016.12.009

Specifically, VR affords investigation of distant locations, exploration of hidden phenomena, and manipulation of otherwise immutable structures (Lee & Wong, 2014). For example, VR can help medical students explore delicate internal organs that would otherwise require cadaver dissection (Nicholson, Chalk, Funnell, & Daniel, 2006). While once a rarity, VR systems are an increasingly commonplace consumer product that may be adopted for instructional use. Yet, currently available, low-cost consumer products typically facilitate observation of virtual environments (e.g., by moving the direction of one's head), support for direct manipulation of structures in the environment is often lacking (Millar, 2016). Without support for direct manipulation, will these systems be effective educational tools? In this manuscript, we explore the role that direct manipulation plays in three-dimensional virtual reality systems by comparing participants who directly manipulate an anatomical structure in a 3-D VR program to those who only view the structure in the same program.

We chose to investigate VR in the context of medical education, because VR programs have the potential to induce the most dramatic shift in anatomy instruction since Vesalius introduced richly illustrated volumes of the human body based on careful, intricate cadaver dissections (Dyer & Thorndike, 2000). While computer technology has undoubtedly transformed the manner in which doctors evaluate and treat their patients, the methods used to teach medical students have been in place for centuries. In particular, cadaver dissection has been considered the gold standard in anatomy instruction dating back to the Renaissance (Dyer & Thorndike, 2000). Although dissection provides students with both a clear view of human organs and their spatial orientation within the body (McLachlan, Bligh, Bradley, & Searle, 2004), the high cost of cadavers and equipment (Robison, Liu, & Apuzzo, 2011; Seymour et al., 2002), the stress placed on medical students (Charlton & Smith, 2000; Finkelstein & Mathers, 1990), and instructional ineffectiveness for small or delicate organs (Hu et al., 2010; Nicholson et al., 2006) present clear limitations. Substituting human cadavers for animals also present ethical challenges and should be minimized (Russell & Burch, 1959; Tannenbaum & Bennett, 2015).

VR represents a promising alternative to cadaver dissection for learning anatomy and practicing surgical procedures (Lee & Wong, 2014). VR systems enable direct interaction with three-dimensional models of anatomical structures. Relative to cadaver dissection, maintenance of a virtual reality 3-D computer model is more cost-effective (after initial development) and sustainable. Likewise, by modeling common physiological processes, such as cancer growth (Jeanquartier, Jean-Quartier, Cemernek, & Holzinger, 2016), computer models have been used to reduce, refine, and replace animal experimentation for biomedical research. Addressing students' comfort working with cadavers, the attitudes of medical school students seem to favor computer models that minimize undue stress (Cabral & Barbosa, 2005; Hariri, Rawn, Srivastava, Youngblood, & Ladd, 2004; Kerfoot, Masser, & Hafler, 2005). Additionally, using virtual 3-D models facilitates magnification of smaller, more delicate structures (e.g., the inner ear) for detailed observation without the physical constraints of cadavers that restrict learner interaction (Nicholson et al., 2006).

Yet, beyond simply providing exposure to relatively inaccessible structures, the potential effectiveness of VR may depend upon the manner in which learners interact with and manipulate represented structures (Lemole, Banerjee, Luciano, Neckrysh, & Charbel, 2007). Specifically, learning anatomy typically requires students to view structures from multiple perspectives, coordinate adjacent structures, and integrate structures into a comprehensive (and potentially hidden) whole (McLachlan et al., 2004). These tasks are highly demanding of spatial cognitive resources (Stull, Hegarty, & Mayer, 2009). Directly manipulating structures in a virtual environment may promote development of “embodied”, multi-modal mental representations of represented structures (Barsalou, 1999). Embodied learning prepares students to engage in mental imagery or simulations in the absence of the physical structures (Barsalou, 1999). For medical students, the ability to imagine and mentally manipulate anatomical structures is a crucial skill (Stull et al., 2009).

In contrast to the embodied view, in which direct manipulation is necessary for learning, it may be that learning anatomy is primarily a function of exposure to optimal information. In this case, video or even still images may be sufficient. Indeed, previous research (e.g., Keehner, Hegarty, Cohen, Khooshabeh, & Montello, 2008) supports an information-processing perspective that de-emphasizes the role of direct manipulation. In the following we survey relevant embodied and information-processing research to explore what features of virtual reality are most likely to promote learning.

1.1. *Embodied cognition and mental imagery*

Standard information-processing theories of cognition view perceptual and motor systems as peripheral to cognition, whereas an embodied view of cognition places elevated significance on these systems (Barsalou, 2008; Clark, 1999; Wilson, 2002). Mounting evidence suggests that what were previously thought to be “purely cognitive” tasks necessarily recruit both perceptual and motor systems. Some of the earliest and most compelling evidence of this comes from the study of mental imagery, which plays a particularly central role in anatomy instruction and medical education (Stull et al., 2009).

Research on mental imagery and rotation has shown that individuals manipulate mental representations much like they would actual objects in physical space, such that the time it takes to mentally rotate an image increases linearly with the degree of rotation (Shepard & Cooper, 1982; Shepard & Metzler, 1971). This research suggests that mental representations not only have perceptual qualities, but that they recruit processes from the motor system, as well (Wexler, Kosslyn, & Berthoz, 1998; Wohlschlager & Wohlschlager, 1998). Neuroimaging studies show that motor cortices (primary/M1 or premotor cortex) are activated when performing mental transformation tasks (Cohen et al., 1996; Kosslyn, DiGirolamo, Thompson, & Alpert, 1998), and that transcranial magnetic stimulation targeted to interfere with neuronal processes in motor regions of cortex reduce mental rotation performance (Ganis, Keenan, Kosslyn, & Pascual-Leone, 2000).

Additionally, the relationship between manual activity and mental rotation can be impacted by context, prior experience, and even voluntary perspectives or strategies applied by the individual. In a PET imaging study of mental rotation, Kosslyn, Thompson, and Wraga (2001) found that participants who had previously rotated an object by hand, rather than passively observed the object being rotated by a motor, showed stronger activity in motor cortex when asked to imagine displayed 3-D blocks being rotated in the same manner as the physical object. According to Kosslyn and colleagues, individuals may voluntarily take alternative perspectives during mental rotation, which may have differential impact on performance.

In particular, objects that are perceived to be anatomical in nature may prime a more embodied approach to mental imagery than abstract figures—i.e., one may project sensorimotor processes to the structure. For example, Armel and Ramachandran (2003) demonstrated that participants, in carefully controlled conditions, were capable of experiencing illusory sensations, such as pain, that corresponded to perceived manipulations performed on an artificial hand. Additionally, anatomical structures promote stronger performance in mental rotation tasks than structures that are perceived to be non-anatomical (Amorim, Isableu, & Jarraya, 2006).

Furthermore, the biomechanical constraints that would affect physical motion appear to influence cognition. For example, Amorim et al. (2006) found that performance advantages for anatomical objects dissappeared when the structure was manipulated into anatomically impossible positions. Likewise, individuals are faster and more accurate at identifying rotated images of, or performing mental rotations on, drawings of hands when the rotations correspond to physically accessible movements (Parsons, 1987a, 1987b; Parsons et al., 1995; Schwoebel, Friedman, Duda, & Coslett, 2001). Finally, mental rotation of an arm-like structure is impeded when impossible positions of the elbow are modeled; although, no such difficulty arises for a hammer-like system (Petit, Pegna, Mayer, & Hauert, 2003).

What mechanisms account for these results that are particular to the mental manipulation of anatomical structures? According to Amorim et al. (2006) individuals are inclined to map anatomical structures to their own bodies' coordinate systems. Similarly, Armel and Ramachandran (2003) claim that the sensory illusions emerged when the participants assimilated the perceived object into their own body image. In particular, the projection of biomechanical constraints on extrinsic objects suggests that individuals are capable of sensing the relative position of the external object to one's own body.

While these studies suggest that anatomical structures typically activate an embodied approach to performance, they make no claim about the impact of embodiment on learning. Recent studies of embodied learning environments suggest that direct manipulation of external representations (physical or virtual) of materials can enhance learning (Black, Segal, Vitale, & Fadjo, 2012; Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004). This is particularly the case when the actions performed are congruent with spatial features of the target concept or structure. For example, instruction promoting grouping gestures (e.g., circling) enhances performance on addition-related tasks (Broaders et al., 2007); linear movements on a board game enhance number line concepts (Siegler & Ramani, 2009); linear movements on a virtual slider enhance understanding of linear forces in science instruction (Chan & Black, 2006, pp. 64–70; Day & Goldstone, 2011); circular movements on a force-feedback joystick enhance understanding of gear-related concepts (Han & Black, 2011).

In these cases, coordination between motion and visual features of the target structure aids learning. For anatomical structures, direct manipulation in VR may promote alignment between the perceived structure and one's own body to facilitate learning. Conversely, surprising movements during passive observation may break this link, and require re-orientation. These individuals must work harder to maintain strong embodiment, which may have a negative impact on learning.

In the case of passive viewing, without a direct link between one's own body and the target structure, individuals may need to depend more heavily on spatial ability. Luursema, Verwey, Kommers, Geelkerken, and Vos (2006) found that the combination of dynamic exploration and stereoscopic display was the most beneficial for low spatial ability participants. Likewise, Meijer and van den Broek (2010) found that active exploration improved low spatial participants' 3-D mental representations of complex 3-D objects, whereas active exploration had no clear effect on middle or high spatial participants' representations. Finally, Lee and Wong (2014) found that lower spatial ability students were more likely to benefit from a desktop VR system when studying anatomy than high spatial ability students. Therefore, by reducing the spatial demands VR may have an added benefit for low visuospatial learners.

1.2. Optimizing information in anatomy instruction

While the embodied cognition framework presented above presents a clear case for computer simulations in anatomy education, empirical investigations have produced mixed findings. Holzinger, Kickmeier-Rust, Wassertheurer, and Hessinger (2009), found equivalent learning outcomes for medical students studying blood dynamics using either a virtual simulation or text, although they found benefits for the simulation with structured guidance. Nicholson et al. (2006) found that an interactive 3-D model of the inner ear facilitated stronger learning outcomes than a series of 2-D images. In this case is not clear whether the improved performance was due to the interactive setting or the 3-D nature of the model.

To address these confounds, Garg and colleagues conducted a series of closely controlled studies concerning interactivity and the importance of accessing certain views (Garg, Norman, & Sperotable, 2001; Garg, Norman, Eva, Spero, & Sharan, 2002; Garg, Norman, Spero, & Maheshwari, 1999). In contrast to Nicholson et al. (2006), they hypothesized that complex 3-D anatomical structures are learned through key viewpoints rather than continuous 3-D orientations of the object. In an initials study, Garg et al. (1999) found no instructional advantage for presenting an anatomical structure (the carpal bone) in multiple successive views at 15-degree over successive presentation of three key views.

In a follow-up study Garg et al. (2001) once again presented participants with either key or multiple views of the carpal bone; however, this time the participants in both groups controlled the transitions to new views. Here multiple view participants performed better than those in the key view group. Learners in the multiple view group spent most of their time on key views, with a notable variation around the 0° and 180° presentations. Although these findings suggest an advantage for active, user-controlled learning of multiple views, Garg and colleagues argued that participants extracted most of the information from the key views with a small amount of "wiggle room", defined as ± 10° around the key views, to gain a sense of the third dimension.

In a third study, Garg et al. (2002) set to reconcile the contrasting findings of their earlier studies. In this study, one set of participants was afforded unconstrained interactivity, while the other group was restricted to motion ± 10° around the key views. When accounting for spatial ability, no difference was found in the learning between groups. Garg et al. (2002) concluded that providing learners with a dynamic, 3-D computer presentation of an anatomical structure provides minimal if any advantages.

Similarly, Keehner et al. (2008) investigated the issue of interactivity and key views. For the first experiment, participants in an interactive group could rotate the structure along the horizontal or vertical axes, whereas the non-interactive group of participants watched a visualization of the same structure rotating repeatedly through alternating horizontal and vertical axes. On a test of spatial inference (i.e., draw the expected cross-section from a specific orientation), participants in the interactive condition performed better than those in the non-interactive condition.

Yet, participants in the interactive condition appeared to stop at key views, while those in the observation condition watched a continuous rotation of the structure, Keehner et al. (2008) thus conducted a second experiment wherein the visual information between conditions was equalized (i.e., a yoked-pairs design). In this case, no benefit for interactivity emerged. Furthermore, in a third experiment, non-interactive participants who observed an optimal set of movements performed as well as interactive participants who spontaneously performed optimal movements, and better than interactive participants who performed sub-optimal movements. Optimal movements, which were based on high-performing interactive participants from the second experiment, emphasized a key view with repeated "wobble" around this position (within 45°), presumably to gather more 3-D information.

The collective findings from Garg, Keehner and their colleagues clearly promote a more efficient, key views approach than advocated by the embodied perspective. Also, these studies did not replicate Luursema et al.'s., (2006) finding that interactivity assists low spatial ability learners, relative to high spatial ability learners. In particular Garg et al. (1999) found that presenting multiple views to low spatial ability participants significantly handicapped their learning of the anatomy. Additionally, both Keehner et al. (2008) and Huk (2006) found that the low spatial ability participants generally had more difficulty with unconstrained 3-D models. These findings led the authors to conclude that 3-D representations should be used carefully with low spatial ability participants, as more views of the target structure may overwhelm these students.

1.3. This study

The studies of learning with virtual, 3-D models presented above provide an inconsistent narrative about how direct manipulation affects learning and how this effect may be moderated by spatial ability. In particular, why did contrasting findings for low spatial ability participants emerge in Luursema et al. (2006) and Garg et al. (1999)? Considering that the medical community is continuing to adopt these programs, there is an immediate need to assess whether students with low spatial ability, who may already be at a disadvantage, will be assisted or harmed by this technology.

It may be the case that the benefit for all participants in the interactive 3-D condition in both Luursema et al. (2006) and Nicholson et al. (2006) – and for low spatial ability participants, particularly, in the former – stems from the confounded comparison to non-interactive, 2-D materials. Evidence from Keehner et al. (2008) would suggest that while the 3-D models may have been beneficial, the interactivity, per se, did not drive learning.

On the other hand, it may be the case that the lack of benefits for interactive conditions found by Garg et al. (1999, 2001) and Keehner et al. (2008) may be rooted in particular qualities of the learning materials employed. In particular, the focus on wrist bones, which are partially visible, allowed participants to make use of their own hand as a supporting 3-D model, even when only key views were presented. Conversely, Keehner's use of non-anatomical structures may not have sufficiently primed an embodied approach to the task. In either case, the potential benefits of direct manipulation in VR may have been mitigated by these alternative considerations. Furthermore, in neither case were stereoscopic models presented to learners, which may further enhance the sense of embodiment with the learning materials.

In the following experiment, we investigate these issues by employing a yoked pair design to compare participants, with varying levels of spatial ability, who actively manipulated a stereoscopic 3-D model to participants who passively observed the movement of these stereo 3-D models, as guided by their yoked partner. We chose to target the inner ear because it is a critical anatomical structure (incorporating a facial nerve that may cause paralysis if damaged), complex, small, and completely hidden from external observation (Nicholson et al., 2006). Therefore, addressing this anatomical structure in an immersive stereo 3-D environment represents a critical test for this technology.

Given supporting theory from the field of embodied cognition, we predict that learners who actively manipulate the anatomical structure will maintain stronger coordination between the model and their own body, which will promote stronger learning than those who passively view the model. Further, we hypothesize that VR will particularly benefit participants with low spatial ability, who may have greater difficulty maintaining embodiment when passively viewing a complex anatomical system. To investigate this possibility, we analyze the motion characteristics of movie clips generated by the interactive participants.

2. Method

2.1. Participants

Seventy-six medical students at a medical school in a metropolitan area of the North East United States participated in this study. Among the seventy-six students, forty-one (54%) were in their first year of medical school, twenty-seven (35%) in their second year, two (3%) in their third year, and six (8%) in their fourth year. Forty-four participants were male (58%), thirty-two participants (42%) female. The ages of the participants ranged from 20 to 38 years, with the majority (65%) between the ages of 22–24 years. Participants were recruited through announcements made after a first-year anatomy class as well as through word of mouth. Medical school students were targeted (rather than residents) for this study because they have not had any formal anatomy instruction of the inner ear. None of the participants had prior exposure to the VR machine.

2.2. Materials

2.2.1. The VR system

The VR machine, Dextroscope® (Volume Interactions, Singapore), is housed at the medical school. RadioDexter, the software that is run within the Dextroscope®, is a medical imaging visualization program that generates a stereoscopic 3-D environment, viewable through stereoscopic glasses. To use the equipment a user is positioned at a kiosk with a mirrored monitor display (left side of Fig. 1). The monitor displays the computer model in two-dimensions. Upon placing stereoscopic glasses over his or her eyes, the computer model is perceived in three-dimensions. To control the computer model the user grips a free-moving (non-mounted) handle ("joystick") that is tethered to a computer through a cable. Users may move or rotate the joystick in any direction within the length of the cable. When running RadioDexter, a user can manipulate the virtual model by deploying a "trigger" button, and then moving the joystick as one would move the target object if holding it. For example, the user can rotate the model by turning the joystick, pan to the right by moving the joystick to the left, and zoom out by moving the handle away from his or her body. Rotation and movement of the virtual structure is only limited by the user's own biomechanical constraints.

2.2.2. Target anatomical structure

A virtual model of the inner ear was created for use in the Dextroscope®. The virtual ear model was created by a licensed otolaryngologist in conjunction with the engineers of the RadioDexter program to ensure accuracy of the ear model (Fig. 2).

2.2.3. Pretest measures

To measure participants' familiarity with inner ear anatomy we administered a four item questionnaire, which we developed in conjunction with a certified otolaryngologist. To measure spatial ability we administered the Vandenberg and Kuse (1978) Mental Rotation Test (MRT), which assesses ability to mentally rotate and visualize a 3-D structure. We also administered Ekstrom, French, Harman, and Dermen's (1976) Building Memory Test of spatial ability (BMT) and a background questionnaire focusing on video game and 3-D model experience; however, the final item was not included in the analysis to limit the scope of the investigation to independently validated covariate measures of spatial ability, which are known to relate to performance within virtual reality.

2.2.4. Posttest measures

Drawing on the work of Keehner et al. (2008), in which participants were asked to draw inferred cross-sections of a structure from several different orientations, participants in this study performed a drawing posttest task. According to these researchers, drawing from multiple viewing angles provides a comprehensive insight into the participant's mental representation of the 3-D structure and represents skills that are valuable in the medical field.

Similar to Keehner et al. (2008), we asked students to draw structures from multiple perspectives; however, to provide a more authentic task we focused on drawing specific sub-structures that are critical components of the system, rather than inferred cross sections. Specifically, participants were provided with successive 2-D images of the inner ear model across five orthogonal planes produced by RadioDexter (lateral, superior, inferior, anterior, and posterior). For each of the five planes, two images were produced in which a key sub-structure was missing (either the facial nerve or the semi-circular canals). Participants were asked to draw a 2-D representation of the missing sub-structure in the correct position and orientation in these ten images.

These sub-parts were specifically chosen due to the complex nature of the shape of each structure as well as the importance each structure plays from a surgical standpoint. In particular, the facial nerve runs parallel to the brain stem and weaves through the center of the inner ear. If the facial nerve is cut, this may cause permanent paralysis to the patient's face. The semi-circular canals are important in helping to maintain a sense of balance, detect movement, and static position. The three semi-circular canals lie perpendicular to each other, one to sense movement in each of the three spatial planes.

Fig. 1. Image of the Dextroscope® virtual reality system, by Volume Interactions. The left-hand side shows the interface for virtual reality users, including a handle (“joystick”). The right-hand side shows the interface for experimenter or teacher.

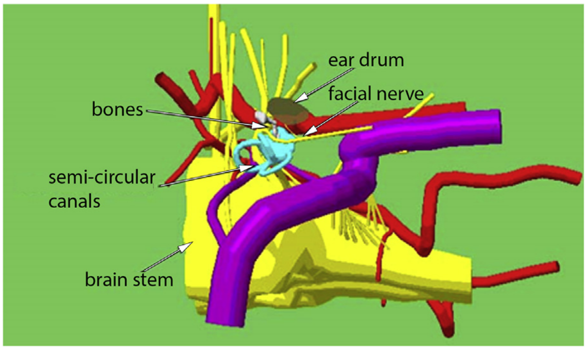


Fig. 2. Screenshot of the ear model from the VR program with labels from surgical plane (between lateral and posterior angles).

2.3. Procedure

Pairs of participants were scheduled to participate in the study during the same session. After meeting the principal investigator and providing consent, each participant completed all four pre-test measures described in the section above. Participants were then randomly assigned to one of the two conditions (*manipulation, viewing*). In the *manipulation* condition participants began training immediately, while in the *viewing* condition participants waited approximately 10 min in order to give the other participant the opportunity to complete his or her training session.

In the *manipulation* condition only, the participant was given a brief training period using the joystick in the VR machine by rotating a CT scan of a human's skull and spine in the system. Once the participant indicated that he or she felt comfortable using the joystick, the target anatomical structure of the inner ear model (Fig. 2) was presented. These participants were informed that they could use the joystick to rotate the ear model as much or as little as they wanted to, and were given 5 min to study. All actions in the *manipulation* session were recorded to allow for "playback" in the 3-D viewing environment.

Upon completing the session, the *manipulation* condition participants were brought to a separate area for the posttest. Subsequently, the paired, *viewing* participant watched the recording of the paired *manipulation* participant's movements in 3-D in the VR machine. This yoked-pair design ensured that both participants viewed exactly the same information. In the *manipulation* condition participants were not told that their interactive session would be viewed by another participant. Likewise, in the *viewing* condition participants were not told that the model they viewed originated from another study participant.

In both conditions the primary model was introduced to participants with a brief explanation of the inner ear, including its initial orientation in the surgical position (i.e., "patient" is lying on the left side of her body, nose pointing forward and right ear facing the ceiling). Participants were asked to study the physical and spatial configuration of the facial nerve and the semi-circular canals. Participants were told that they would be asked to draw these structures from multiple perspectives at posttest.

For the ten posttest items participant was asked to draw, to the best of his or her ability the missing sub-structure on each of ten images. Participants were not timed when completing this post-test. The order of the images given to the participants was randomized (by missing sub-structure) to eliminate any learning effect. For the images of the missing semi-circular canals, participants were asked to write on the paper how far apart they were trying to draw each canal from the other.[1]

2.4. *Coding and analysis*

2.4.1. *Posttest*

The drawings of the facial nerve were assessed for accuracy of visual representation on three criteria: parts, angle and placement. The drawings of the missing semi-circular canals were assessed for accuracy of mental visual representation on four criteria: parts, angle, placement and size. These codes were applied to all the ten images, with changes in the specific criteria reflecting the particular orientation (anatomical plane) of the structure.

The researcher and an independent coder coded the post-test. The second independent coder was trained using a random subset of the data to gain experience with the coding scheme. Both coders were blind to the identity of the participants and condition assignment each coded all 760 drawings, the overall percent agreement between researchers was 95%. The researcher then reviewed the scores and identified any disagreements. These disagreements were resolved by discussion between the two coders. Posttest subtotals were then computed by anatomical structure and anatomical plane, as well as an overall total (TOTAL).

2.4.2. *Video analysis*

From any interactive session the RadioDexter outputs a 2-D movie clip that displays the animation of the anatomical structure as it was viewed by a user. All clips are 640 pixels wide and 480 pixels tall. No other numerical information was provided to the researchers. In lieu of precise data about users' viewing perspective and actions performed within a training session we utilized video analysis software to summarize motion dynamics. Specifically, the open-source software program, *Tracker: Video analysis and modeling tool* (Brown, 2009), heretofore *Tracker,* is part of a larger open source project called the *Open Source Physics Project* whose goal is to provide high quality, free software for science education.

While *Tracker's* primary purpose is to track common moving objects in videos for physics instruction (e.g. the movement of a ball), the particular analysis algorithm of this software was well-suited to the anatomical videos. Specifically, the software automatically tracks selected objects in videos by matching similar color patterns across frames. In this case, because the bone structure of the middle ear (ossicles) produces a relatively small, distinct image with a unique hue, *Tracker's* algorithm was capable of following a continuous path of this structure.

For each video this inner ear bone structure was tracked across both the x- and y-axes over alternating frames of the movie clip, at 0.08 s interval (Fig. 3). The tracking procedure was semi-automated, such that tracking continued until the target structure had moved out of the expected range (based on previous location and velocity), typically due to a sudden movement by the participant. In these cases, the experimenter manually reset the tracked point back on the target structure and resumed automated tracking. In cases where the bone structure was not in view (i.e., off-screen or obscured by the brain stem) the frame was simply left unmarked. Because the target structure was not perfectly regular (in shape or color) it was impossible to maintain tracking at the exact center of the object throughout the entire clip. Therefore, some degree of noise was introduced into the video artifact (typically measuring no more than 20 screen pixels). To account for this noise, we applied median and LOESS regression to smooth the data.

Because only a single point was tracked, the viewing perspective of the participant could not be derived, as the x-y co-ordinates of the ossicles could be produced from any number of perspectives. However, the change in perspective was

---

[1] Based on the results of a pilot study, it was determined that drawing the angular relationship among the semi-circular canals is difficult on paper; therefore, we decided that having participants write how far apart they were drawing each canal from the other would not penalize those participants who could not draw what they were visualizing.

generally accompanied by a dramatic shift in x-y coordinates of the tracked point. On the other hand, participants engaging in a "wiggle" motion – i.e., fine movements within a narrow visual perspective – produced much less, but non-zero, changes in x-y coordinates. Therefore, the supplied x-y data was appropriate for analyzing the magnitude of motion, rather than exposure to particular key views.

To summarize these interaction dynamics, we calculated a primary measure of dynamism – *mean displacement per second* (MDps) – by summing the x-y screen pixels traversed in the study session, using Pythagorean distance, and dividing by number of seconds. In addition, we further distinguished motion by degree of magnitude. As described previously there may be two qualitatively different approaches to movement: "wiggle" to derive more information within a view and "shift" to change perspectives. To represent this numerically we automatically searched clips for short intervals (1.7 s) with displacements greater than threshold (66.3 pixels) to tag as "major motion". We then tagged all other short intervals with greater than a much a smaller threshold (8.5 pixels), to account for residual noise, to tag as "minor motion". These threshold values were chosen by observing a set of four videos (those displayed in Fig. 7, in results), finding instances of major motion, calculating their average time span (1.7 s) and the minimum displacement *per* second (39 pixels, i.e., 66.3 pixels over 1.7 s).

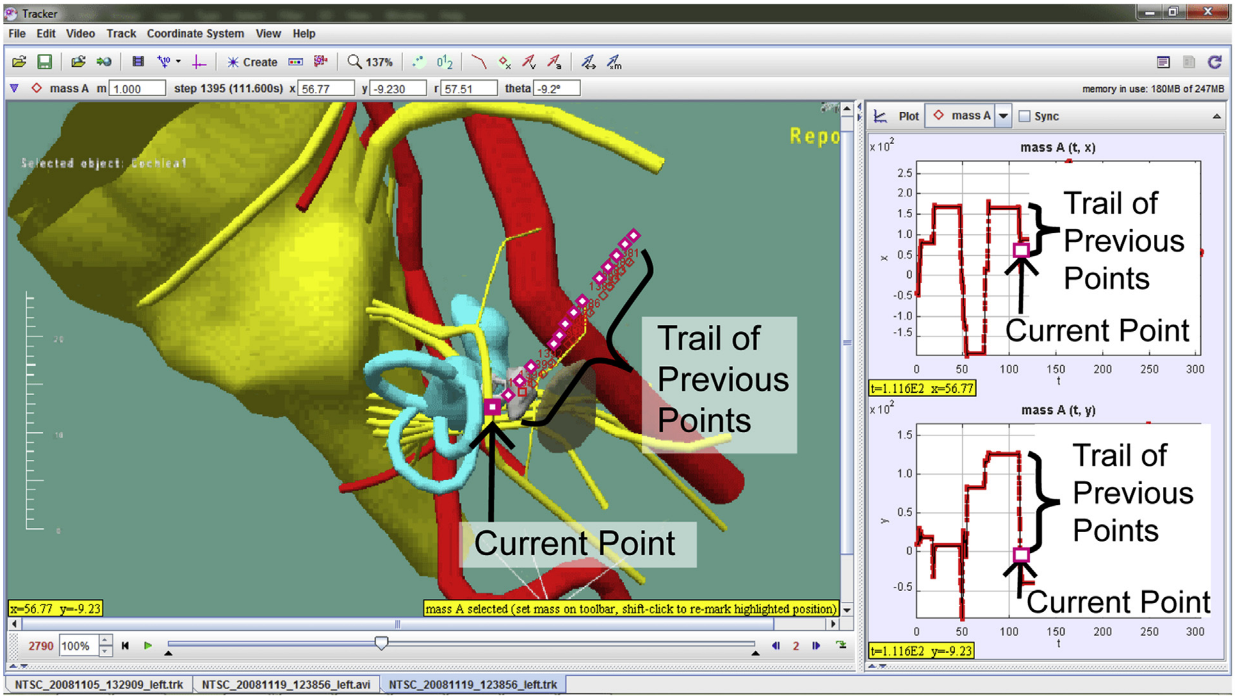
From these tagged intervals we computed two more summary statistics – *percent major motion* (PMajM) and *percent minor motion* (PMinM) – by dividing the total amount of time engaging in motion (major or minor, respectively) by the total duration of the clip. The remaining percent of time can be considered percent motionless.

3. Results

3.1. Comparison of manipulation and viewing conditions

The two groups of participants, differentiated by experimental condition, did not differ in either pretest ear anatomy knowledge [*manipulation:* M = 2.0 (SD = 0.9); *viewing:* M = 2.0 (SD = 0.9); t (37) = 0, p = 1], nor spatial ability as assessed by raw MRT score [*manipulation:* M = 18.6 (SD = 8.1); *viewing:* M = 19.8 (SD = 7.8); t (37) = −0.7, p > 0.1] or raw BM score [*manipulation:* M = 21.1 (SD = 2.6); *viewing:* M = 21.7 (SD = 2.7); t (37) = −1.0, p > 0.1].

In confirmation of our main hypothesis a *t*-test of paired posttest TOTAL revealed that participants from the *manipulation* condition achieved significantly higher posttest TOTAL scores than their yoked partners in the *viewing* condition [*manipulation:* M = 72.5 (SD = 7.3); *viewing:* M = 60.8 (SD = 11.4); t (37) = 6.44, p < 0.000, d = 1.04]. The large effect size, i.e., greater than one standard deviation, indicates a strong influence of experimental condition.



mass A m 1.000 step 1395 (111.600s) x 56.77 y -9.230 r 57.51 theta -9.2°

memory in use: 180MB of 247MB

x=56.77 y=-9.23

mass A selected (set mass on toolbar, shift-click to re-mark highlighted position)

t=1.116E2 x=56.77

t=1.116E2 y=-9.23

Fig. 3. Annotated image of *Tracker* software. Main window (left) displays anatomical structure, including the pale gray bone structure being tracked. A series of (color-enhanced) points displays the previous location of the tracked object. Corresponding information is found in graphical form on the right. The top graph displays x-position over time, while the bottom graph displays y-position over time. The participant (S1) is currently engaging in a rotation in both x- and y-dimensions.

In addition to the combined TOTAL score, we further differentiated scores by both anatomical sub-structure and anatomical viewing plane (Fig. 4). For TOTAL scores differentiated by anatomical sub-structure repeated-measures ANOVA revealed a significant effect of condition [$F(1, 37) = 46.0$, $p < 0.001$, $\eta_p^2 = 0.55$], a significant effect of target sub-structure [$F(1, 37) = 84.0$, $p < 0.001$, $\eta_p^2 = 0.69$], but no significant interaction between condition and anatomical sub-structure [$F(1, 37) = 0.46$, $p > 0.1$, $\eta_p^2 < 0.01$]. While participants drew the semi-circular canal more accurately than the facial nerve, this difference was consistent across conditions.

Likewise, with TOTAL scores differentiated by anatomical viewing plane repeated-measures ANOVA revealed a significant effect of condition [$F(1, 37) = 46.0$, $p < 0.001$, $\eta_p^2 = 0.55$] and viewing plane [$F(4, 148) = 20.0$, $p < 0.001$, $\eta_p^2 = 0.35$]. Additionally, unlike sub-structure, a significant, but small interaction between viewing plane and condition did emerge [$F(4, 148) = 2.8$, $p < 0.05$, $\eta_p^2 = 0.06$]. While the reason for this interaction is not immediately clear, the viewing plane with the least difference between conditions – the lateral plane – was the closest to the introductory (surgical) position of the model, and therefore likely received the most exposure by both conditions.

3.2. *Individual factors – prior knowledge and spatial ability*

Overall, prior knowledge of ear anatomy scores (EA) were not significantly correlated with posttest TOTAL scores [$r = 0.19$, $t(74) = 1.6$, $p > 0.1$]; however differentiated by condition EA scores were correlated with TOTAL scores in the *manipulation* condition [$r = 0.39$, $t(36) = 2.6$, $p < 0.05$], but not in the *viewing* condition [$r = 0.11$, $t(36) = 0.68$, $p > 0.1$].

To further investigate these relationships we performed a regression on TOTAL, with condition (dummy-coded: 1 for *viewing*, 0 for *manipulation*), EA, and EA × condition as predictors. In this model neither condition nor EA × condition were significantly associated with TOTAL [condition: $B = -8.4$, $SE_B = 5.3$, $\beta = -0.38$, $t(72) = -1.6$; $p > 0.1$; EA x condition: $B = -1.65$, $SE_B = 2.4$, $\beta = -0.18$, $t(72) = -0.7$; $p > 0.1$]; however, EA (i.e., EA for *manipulation* participants) showed a trend towards significance [$B = 3.05$, $SE_B = 1.6$, $\beta = 0.25$, $t(72) = 1.9$; $p = 0.07$]. Fig. 5 demonstrates a stronger association between EA and TOTAL in the *manipulation* condition.

Of the two tests used to measure spatial ability – i.e., the Ekstrom et al.'s Building Memory Test (BMT) and Vandenberg & Kuse's Mental Rotation Test (MRT) – only MRT demonstrated a significant correlation with posttest TOTAL [$r(74) = 0.33$, $p < 0.005$]. Therefore, only MRT is used in further analyses of spatial ability.

Looking more closely at the role of spatial ability on the two different learning modalities, correlations differentiated by condition revealed a significant association between TOTAL score and spatial ability for the *viewing* condition [$r(36) = 0.58$, $p < 0.001$], but no such correlation for the *manipulation* condition [$r(36) = 0.23$, $p > 0.1$].

Likewise, we performed a regression on TOTAL, with condition, MRT, and MRT × condition as predictors. In this model, in contrast to the model with EA, significant effects emerged for both condition [$B = -24.7$, $SE_B = 5.07$, $\beta = -1.1$, $t(72) = -4.9$; $p < 0.000$] and MRT × condition [$B = 0.64$, $SE_B = 0.24$, $\beta = 0.65$, $t(72) = 2.6$; $p < 0.05$]; however, MRT (i.e., MRT for *manipulation* participants) was not significantly associated with TOTAL [$B = 0.21$, $SE_B = 0.17$, $\beta = 0.15$, $t(72) = 1.2$; $p > 0.1$]. Fig. 6 demonstrates a stronger association between MRT and TOTAL in the *visual* condition.

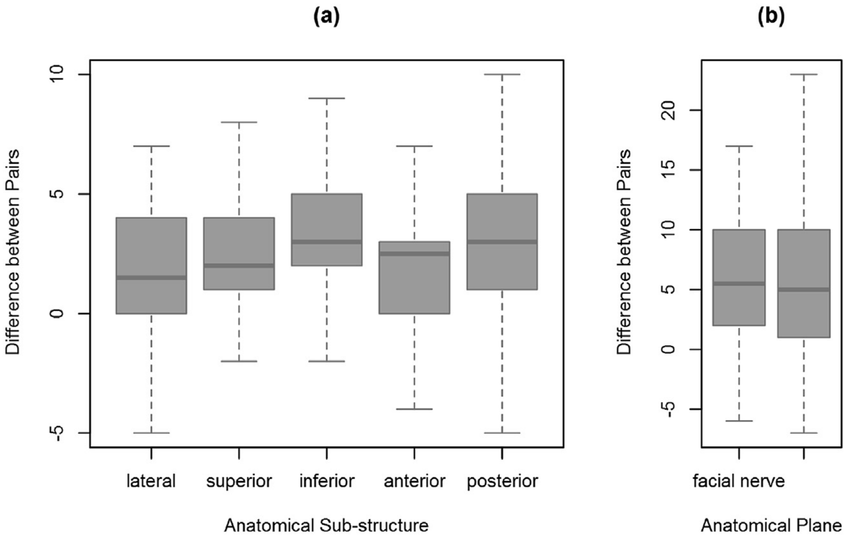


Fig. 4. Box-and-whiskers plot displaying differences between pairs on posttest scores (manipulation – observation), separated by anatomical plane (a) and sub-structure (b). The dark line within each box indicates the median value.

3.3. Video analysis

As described in the method section, we applied video analysis software (*Tracker)* to track the screen coordinates of the ossicles of the middle ear, which are positioned adjacent to the semi-circular canal – a target of posttest drawing. Fig. 7 displays several examples of the x- and y-coordinate data over time produced by *Tracker*. These time-series plots reveal major differences in the dynamic qualities of interactive sessions. For example, the participant in 7c displays nearly constant motion, but very few major shifts of perspective. On the other hand, participant in 7a displays many major shifts in perspective, but appears to remain situated at each perspective for several seconds.

For each of these videos we computed the *mean displacement per second* (MDps) [$M = 22.0$ pixels *per* second ($SD = 9.0$)], the *percent major motion* (PMajM) [$M = 21.9\%$, ($SD = 10.9\%$)], and the *percent minor motion* (PMinM) [$M = 39.4\%$, ($SD = 13.8\%$)]. The sum of these latter motion statistics indicate that participants spent more than 60% of the duration of the clip in motion. Although, on average, participants spent more time engaging in small movements than large shifts of perspective. Specifically, in only 7 of 38 clips did PMajM exceed PMinM (such as Fig. 7a).

As evidence of the validity of this measures we performed a comparison to experimenters' qualitative interpretation. Specifically, two experimenters viewed videos and coded each as *dynamic* or *static* according to the degree to which the participant appeared to remain in motion while observing the model, as opposed to remaining fixated in key views to observe the anatomical model (with some "wiggle room"). We then chose six agreed-upon examples that characteristically represented each category (approximately $^1/_3$ of the pairs). In Table 1, we display the quantitative properties of each of these 12 selected videos. As predicted these two sets of interactions differ significantly in terms of MDps [$t\ (10) = 3.0$, $p < 0.05$, $d = 1.7$].

Furthermore, as can be observed in Fig. 7c, further differentiation into major and minor motion explains why a video categorized as "dynamic" could have the lower value of MDps than all selected "static" videos: the participant engaged in near constant "minor motion". While this individual did change perspectives in the course of the video, the shifts were slow, and therefore did not reach the threshold displacement for "major motion". Interestingly, as Table 1 displays, this was one of the rare cases in which the *viewing* participant outperformed the *manipulation* participant (although the former's spatial ability scores were also higher than the latter's).

To apply these motion statistics beyond the selected examples we first compared participants in the *manipulation* condition who produced highly dynamic videos to those who produced less dynamic videos. Classified according to a median split of *mean displacement per second* (median value = 21.8 pixels), there was no significant difference between groups categorized as high motion or low motion in terms of MRT spatial ability scores [*high motion:* $M = 19.5$ ($SD = 7.6$); *low motion:* 17.8 ($SD = 8.8$); $t\ (36) = 0.6$, $p > 0.1$] or pretest knowledge (EA) scores [*high motion:* $M = 2.2$ ($SD = 0.9$); *low motion:* $M = 1.8$ ($SD = 1.0$); $t\ (36) = 1.2$, $p > 0.1$].

To investigate the relationship between posttest performance and *mean displacement per second* (MDps), we performed a regression on TOTAL, with condition, MRT, MDps, and MDps × condition as predictors. In this model neither condition nor MDps were significantly associated with TOTAL [*condition:* $B = -2.3$, $SE_B = 5.3$, $\beta = -0.10$, $t\ (71) = -0.4$; $p > 0.1$; MDps: $B = 0.24$, $SE_B = 0.16$, $\beta = 0.19$, $t\ (71) = 1.5$; $p > 0.1$]; however, the interaction between condition and MDps (i.e., MDps for *viewing* participants) did reveal a significant negative association with TOTAL [$B = -0.45$, $SE_B = 0.22$, $\beta = -0.52$, $t\ (71) = -2.1$; $p < 0.05$]. MRT scores in this model were also positively associated with TOTAL [$B = 0.54$, $SE_B = 0.13$, $\beta = 0.38$, $t\ (71) = 4.3$; $p < 0.000$]. Fig. 8 demonstrates how the conditions diverge as MDps increases.

We then repeated this regression analysis with both motion parameters, PMajM and PMinM, substituting for MDps, to determine how the magnitude of motion impacted participants in either condition. For PMajM the general pattern of results were similar: neither condition nor PMajM were significantly associated with TOTAL [*condition:* $B = -3.6$, $SE_B = 4.4$, $\beta = -0.16$, $t\ (71) = -0.8$; $p > 0.1$; PMajM: $B = 15.4$, $SE_B = 12.9$, $\beta = 0.15$, $t\ (71) = 1.2$; $p > 0.1$]; however, the interaction between condition

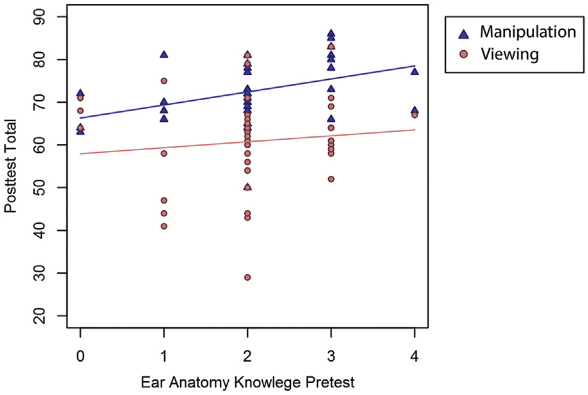


Fig. 5. Differences on posttest TOTAL score performance between the two conditions by pretest knowledge of ear anatomy (EA). Regression lines fit according to following equation: $TOTALi = 66.3 + 3.0 * EAi - 8.3 * Ci - 1.7 * EAi * Ci$ [where C = 1 for observation condition, 0 for manipulation condition].

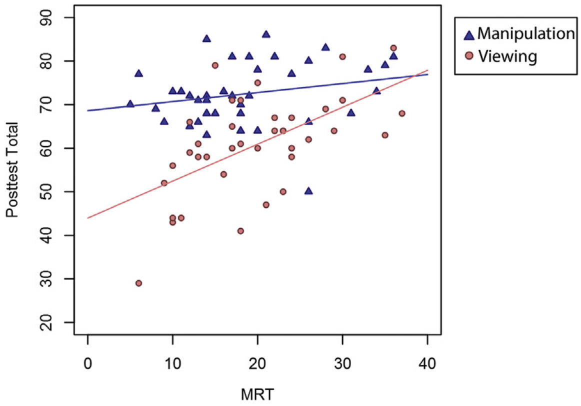


Fig. 6. Differences on posttest TOTAL score performance between the two conditions by spatial ability (Mental Rotation Test). Regression lines fit according to following equation: TOTALi = 66.6 + 0.21 * MRTi − 24.7 * Ci − 0.64 * MRTi * Ci [where C = 1 for observation condition, 0 for manipulation condition].

and PMajM did reveal a significant negative association with TOTAL [$B = -39.7$, $SE_B = 18.2$, $\beta = -0.48$, $t(71) = -2.2$; $p < 0.05$]. MRT scores were also positively associated with TOTAL [$B = 0.51$, $SE_B = 0.13$, $\beta = 0.36$, $t(71) = 4.1$; $p < 0.000$].

On the other hand, a similar analysis with PMinM revealed different relationships. Specifically, like in the previous model the motion parameter PMinM was not associated with TOTAL [$B = -2.4$, $SE_B = 10.3$, $\beta = -0.03$, $t(71) = -0.2$; $p > 0.1$], and also, in this case, the interaction between condition and PMinM was not associated with TOTAL [$B = 22.1$, $SE_B = 14.5$, $\beta = 0.44$, $t(71) = 1.65$; $p > 0.1$]. On the other hand, condition, on its own, was negatively associated with TOTAL [$B = -20.9$, $SE_B = 6.0$, $\beta = -0.94$, $t(71) = -3.5$; $p < 0.001$]. Once again, MRT scores were positively associated with TOTAL [$B = 0.47$, $SE_B = 0.13$, $\beta = 0.33$, $t(71) = 3.6$; $p < 0.001$]. Overall, this model explained a significant proportion of variance [$R^2 = 0.44$, $F(4, 75) = 13.9$, $p < 0.000$].

4. Discussion

4.1. Manipulation vs. observation

The results of this study clearly indicate that, in regards to anatomy instruction, there is added value for directly manipulating virtual 3-D structures beyond simply viewing these structures. Because the posttest was designed to assess the fidelity of participants' internal representation of the spatial features of selected anatomical structures (e.g. shape, location, orientation), we may infer that direct manipulation aided in the process of constructing an internal representation of this structure.

In addition to general learning gains participants in these conditions differed in other ways. Specifically, participants in the *manipulation* showed no significant relationship between spatial ability and posttest outcomes, whereas participants in the *viewing* showed an expected positive relationship between these measures. Inversely, participants in the *manipulation* condition showed a positive relationship between prior knowledge and outcomes, whereas no such relationship emerged in the *viewing* condition. These diverging results suggest that the different modes of presentation resulted in different learning processes.

In the *manipulation* condition the positive effect for prior knowledge suggests that participants with greater background knowledge were more capable of effectively using the system to learn new material. This finding parallels research on learning in more traditional formats (e.g. text reading), where advanced organizers, pre-questions, and predictions prime learners prior knowledge structures, and often lead to stronger learning gains (Pressley et al., 1992). In our study, because participants were aware of learning objectives, those with relevant prior experience may have engaged in more targeted behaviors to address missing knowledge. Similar studies should be conducted that vary participants' prior exposure to target concepts.

On the other hand, a lack of a clear relationship between motion dynamics and learning outcomes for those in the *manipulation* condition suggests that the means of attaining this knowledge were idiosyncratic, perhaps reflecting differences in engagement style, enthusiasm, conceptual background, etc. In other words, the interactive multimedia environment enabled participants in the *manipulation* condition to tailor their experience to their individual cognitive needs and interests (Kirsh, 1997). In contrast, without the ability to proactively direct their experience to fit their own needs, participants in the *viewing* condition were more influenced by immutable factors, such as spatial ability and dynamics of the video.

Yet, why in the *manipulation* condition was the relationship between spatial ability and learning largely absent? Extending our earlier hypothesis, we suspect that the embodied nature of the task helped these participants utilize their own body's reference frame to maintain orientation in the virtual environment. The link between the body and virtual model allowed learners to engage in broad movements, yet still maintain awareness of how any unique view of the model related to the

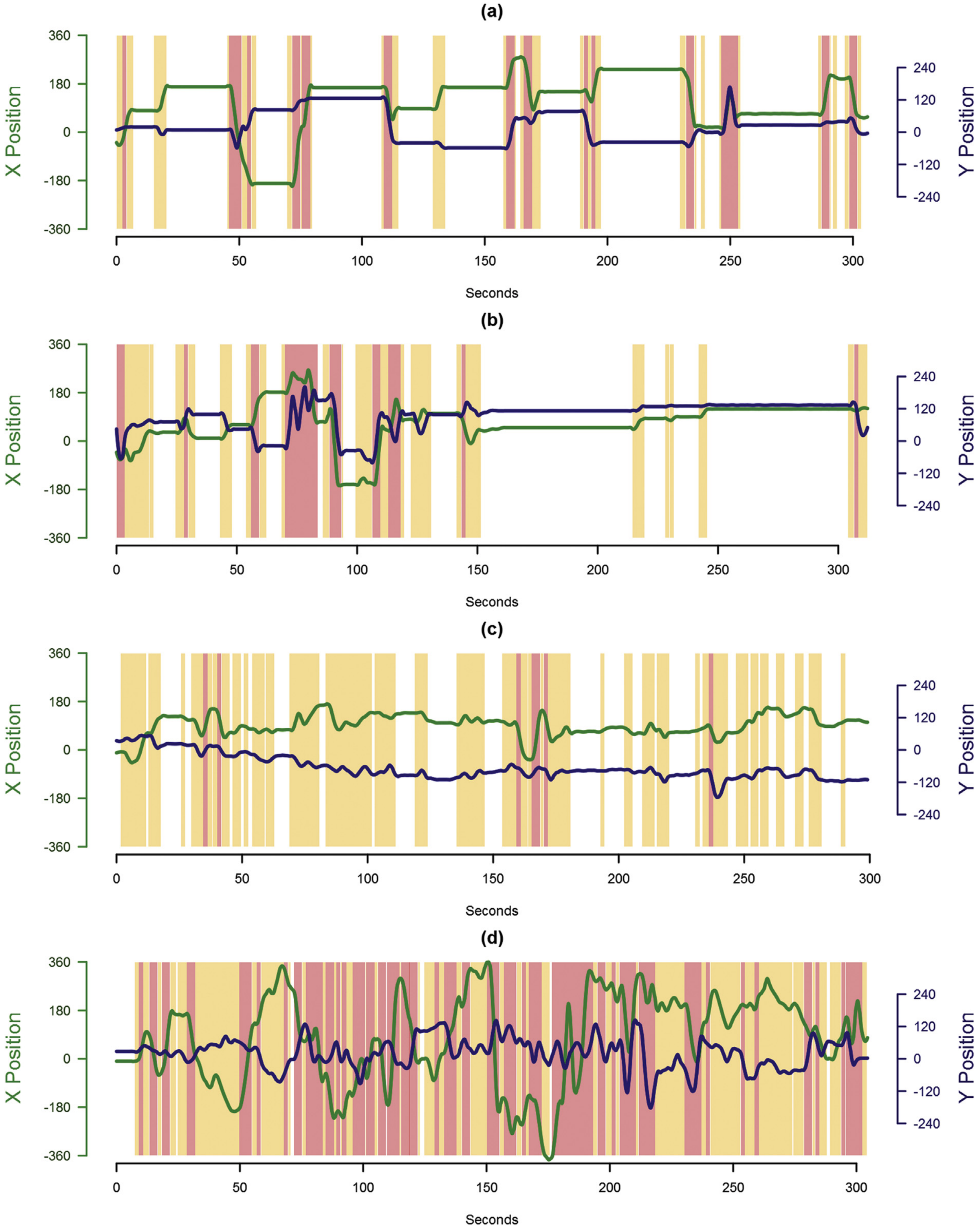


Fig. 7. Screen coordinates of inner ear bone structure over time. Position along x- and y-axes are represented by the green and blue curves, respectively. Orange vertical bars represent intervals of minor displacement, red vertical bars represent major displacement, and remaining white background represents little or no movement. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

Table 1
Comparison between selected highly dynamic and highly static videos.

	Mean displacement per second (pixels)	% Major motion (of total duration)	% Minor motion (of total duration)	Posttest TOTAL score		MRT score	
				Manip.	Obs.	Manip.	Obs.
Static							
S1 (Fig. 6a)	10.7	14.8%	14.3%	78	43	33	10
S2 (Fig. 6b)	11.9	12.4%	15.0%	71	58	14	14
S3	16.8	17.4%	36.9%	73	67	34	24
S4	14.7	14.3%	19.1%	72	60	17	24
S5	12.3	10.1%	48.6%	72	60	12	17
S6	12.2	10.7%	38.1%	68	67	31	22
Dynamic							
D1 (Fig. 6c)	10.1	3.9%	76.3%	65	71	12	30
D2 (Fig. 6d)	42.2	48.6%	29.7%	86	64	21	23
D3	23.3	20.7%	49.4%	81	71	19	17
D4	26.2	30.8%	31.9%	72	47	14	21
D5	26.2	25.8%	31.6%	71	79	13	15
D6	26.6	26.3%	38.9%	70	64	5	29

overall 3-D structure. In fact, the orientation of the VR user's hand as he or she rotated the virtual structure served as a physical trace of the motion. For example, if a participant rotated his or hand 90° yaw, he or she would not find it difficult maintain awareness of how this new (anterior) view related to the default (lateral) perspective because the rotation of the user's wrist would maintain this information. In other words the coordinated relationship between the body and model distributed some of the cognitive load to the mechanics of the body (Zhang & Norman, 1994).

On the other hand, participants in the *viewing* condition could not use their body to maintain coordination across changing perspectives of the structure. Rather, these individuals had to rely upon spatial ability to maintain orientation as the structure rotated through multiple perspectives. If a participant lost track of the orientation of the structure ("am I looking from above or below?") he or she would be unable to integrate the visual information with the overall structure. The demand on spatial ability for these participants was clearly evident by the fact that no participant in the *viewing* condition with a spatial ability lower that one standard deviation from the mean (6 participants with TOTAL of 11.28 or less), achieved a higher TOTAL score than any corresponding participant in the *manipulation* condition (also 6 participants).

4.2. Limitations and future research

4.2.1. The role of embodiment

While the joystick, by virtue of its physical properties ensured that participants in the *manipulation* condition were engaging in motor behavior, which we posit supports learning, we cannot be sure that it was these physical actions, per se, or simply control over the system that facilitated learning. For example, it might be the case that verbally dictating commands (e.g., "move right") would be equally effective as a motor input. Inversely, control over the environment may have played little or no role in comparison to the haptic feedback of the joystick linked to the visual model. In this case a "force-feedback" joystick, which moved independently of the user, could be compared to the use of a user-controlled joystick. While we suspect that both control and motor actions played a role in learning both of these possibilities are open to further research.

While there may be several factors that contributed to the strong performance of participants in the *manipulation* condition, it is not clear what strategies participants in the *viewing* condition used with any moderate success. Given the short time frame of the study, participants in this condition may simply not have had enough time to generate effective viewing strategies. With training and more experience, could participants in a passive condition utilize embodied strategies to help coordinate their body with the model, and therefore enhance learning? In work on gesture research, prompting research participants to make appropriate gestures has been shown to enhance learning (Broaders et al., 2007; Vitale, Swart, & Black, 2014). Could passive viewers learn to use gestures to mimic the behavior of those controlling the model? Considering that most learning experiences, for the foreseeable future, will not be conducted in a VR environment, applying embodied learning strategies to more traditional presentation formats is an important area of research (Black et al., 2012). Video analyses of participants' gestures would be particularly helpful in illuminating the relationship between embodied movement and learning (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001).

4.2.2. Optimal view

These results appear to contrast with earlier work that revealed minimal benefits for direct manipulation. Specifically, no differences were found in a similar study with a yoked pair design, and in a study where passive observers viewed an optimal series of images they performed better than those with direct control (Keehner et al., 2008). In our case, however, we did find differences between yoked partners. Yet, it may be that these differences were driven by a large sample of non-optimal performances. Would our results have dissipated if participants solely viewed an optimal performance? To generate a model of optimal performance we would look for cases where the participants in the *viewing* condition performed as well or

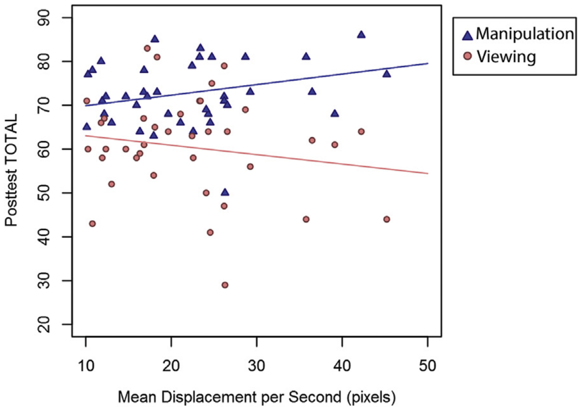


Fig. 8. Posttests TOTAL scores by mean displacement *per* second. With greater movement the conditions diverge. Regression lines printed according to following equation: TOTALi = 57.19 + .54 * MRTi − 2.31 * Ci + 0.24 * MDpsi − 0.46 * Ci * MDpsi [where C = 1 for observation condition, 0 for manipulation condition. Mean MRT (19.2) substituted for all MRTi in this plot].

better than those in the *manipulation* condition. However, this only occurred in 5 out of 38 cases. Furthermore, in each of these 5 pairs the *viewing* participant had a higher spatial ability score than his or her counterpart. Therefore, it is not clear that any one session represents an optimum model.

Moreover, it is the case that for *viewing* participants the amount of motion was inversely related to posttest performance. This dovetails with Keehner's finding that the optimal performance could be described as steady and consistent, with few major shifts of position or haphazard motion. It may be the case that, given the novelty of the virtual reality environment, *manipulation* participants found it difficult to resist engaging in a more expressive, but non-optimal strategies. In this case an optimal solution may be constructed a priori, incorporating only precise, limited transformations.

While this artificial performance could be considered optimal, it may be the case that it is only optimal for the *viewing* condition. In the *manipulation* condition, on the other hand, optimization is dependent upon the individual characteristics of the participant. Achieving the optimal experience in a *manipulation* condition may be a matter of training the user to interact with the system more effectively. The role of training students to utilize virtual reality systems effectively, and the role of transfer of skills between virtual environments, represents an additional avenue of future research.

4.2.3. Video analysis

While the video analysis was applied successfully to establish a negative association between the degree of motion and performance in the *viewing* condition, these interactions contained a great deal more information than we were able to measure with the artifact video clips. In particular, we were unable to establish a link between interaction style and learning outcomes in the *manipulation* condition. While it may be the case that no such link exists, it is possible that additional analyses, with more precisely represented data, could have illuminated a latent relationship. Given the diversity of manipulation styles, a much greater sample of participants would likely be required to establish any effects quantitatively.

Our video analysis, while delivering interpretable results regarding overall motion, also delivered several null results, including a relationship between motion and learning outcomes in the *manipulation* condition or the role of minor motion (i.e., “wiggling”) in either condition. While these may be a true negative results, several features of our analysis approach could have caused us to miss a real relationship. In particular, while a large change in the captured x-y coordinates always indicated a dramatic shift in the movie clip, the converse was not necessarily the case. For example, in instances where the participant rotated around a centered tracked point no change in position would emerge. Likewise, if the participant zoomed in or out from the point, no coordinate position change would emerge. Therefore, our approach likely systematically underrepresented motion, particularly for those who centered their viewing perspective on the ossicles.

While the potential, albeit minor, misrepresentation of motion magnitude reflects a limitation of the particular tracking approach, a more serious limitation is the loss of two dimensions of information, the depth and the orientation of the tracked objects. Without this information we cannot conclude which perspective the user viewed the structure from, and how this evolved over time. This information would be invaluable in determining the degree to which an interaction optimized access to needed information.

Clearly, a more precise numerical representation of movie clips is necessary for this analysis. For virtual models an interaction log can be constructed to track participant actions. For physical models post hoc tracking can be enhanced by using multiple visually-distinguishable markers on the structure to afford computation of the structure's position and orientation. Given sufficient information about the actions performed on the structure we could then begin to assess how participants transitioned between views, “wiggled” within a view, zoomed in and out to focus on local and global features, etc. Analysis on these features of performance may then either reveal general optimal interactions, or how participants utilize the system, based on their own prior knowledge, to create an experience that is optimal for them.

In spite of the limitations of our current approach, we suspect that no single optimal style would emerge from a more precise analysis. While a theoretical approach emphasizing information optimization suggest that systematic and smooth motion, with frequent observational breaks, would facilitate greater learning than more dynamic interactions, many examples contradicted this. For example, Fig. 7d represents a highly chaotic, dynamic interaction. Yet, in this case the participant who manipulated the structure received the highest score within this sample on the posttest (86). Inspection of these videos suggests that there are a wide range of styles, all of which may facilitate learning.

5. Conclusions

The goal of this study was to investigate the impact of direct manipulation in virtual reality on anatomy learning. Building upon prior research showing mixed findings for the role of direct manipulation in spatially-intensive learning environments we attempted to incorporate the strongest features of this research into our study: First, the hand-held VR controller was designed to be highly intuitive and enable movements that were spatially congruent with the actions that would be taken to manipulate a physical model. We suspect that this verisimilitude enabled *manipulation* participant to maintain an embodiment with the anatomical structure. Second, participants, in both conditions, wore stereoscopic 3-D goggles enabling a more realistic visual representation of the anatomical structure; much of the reported literature incorporates monoscopic 3-D or only provides stereoscopic 3-D goggles to treatment participants. Third this study incorporated a realistic anatomical structure, with a specific learning task, rather than a more abstract, “anatomy-like” structure. We suspect that this increased the likelihood that learners spontaneously engaged in embodied processes. Fourth, the study focused on an internal anatomical structure rather than an external structure (such as the wrist), thereby ensuring that participants, in both conditions, could not manipulate the congruent structure on their own body, which could unintentionally mitigate differences between conditions.

The results of the study demonstrate that 1) participants are capable of successfully embodying virtual representations of internal anatomical structures if they can control the presentation; 2) participants who passively view the movement of this structure are most successful when presented with a limited number of canonical viewpoints; and 3) where the VR environment is designed to be intuitive and similar to the physical interaction in the real world, it is the participants with low spatial ability who tend to benefit most from the advantages of manipulation and interactivity, as compared with those with high spatial ability.

These findings suggest that there is a promising future for VR technology in medical education and training. VR and computer models, in general, offer medical students and professionals the opportunity to continue training when access to patients or cadavers are limited (Hessinger, Holzinger, Leitner, & Wassertheurer, 2008). Ongoing projects to construct detailed, complete models of human physiology are likely to pay large dividends for both medical training and research (Hunter et al., 2013). Yet, further research into interactive controls and guidance for these models is still needed. We also encourage greater use of VR in cognitive and psychological research as a means of testing hypotheses regarding embodied cognition.

Acknowledgements

We would like to thank the students and faculty of the University of Medicine and Dentistry of New Jersey (now Rutgers Biomedical and Health Sciences in Newark). Susan would like to thank the members of her dissertation committee for their patience and guidance, including: Herbert Ginsburg, Matthew Johnson, John Zimmerman, and Mark Graham.

References

Amorim, M.-A., Isableu, B., & Jarraya, M. (2006). Embodied spatial transformations: “body analogy” for the mental rotation of objects. *Journal of Experimental Psychology General, 135*(3), 327–347.

Armel, K. C., & Ramachandran, V. S. (2003). Projecting sensations to external objects: Evidence from skin conductance response. *Proceedings. Biological Sciences/The Royal Society, 270*(1523), 1499–1506.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–609.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617–645.

Black, J. B., Segal, A., Vitale, J. M., & Fadjo, C. L. (2012). Embodied cognition and learning environment design. In D. Jonassen, & S. Land (Eds.), *Theoretical foundations of learning environments* (2nd ed., pp. 198–223). New York: Routledge.

Broaders, S. C., Wagner, S. C., Zachary, M., Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General, 136*(4), 539–550.

Brown, D. (2009). *Tracker video analysis and modeling tool*. Retrieved from http://www.cabrillo.edu/~dbrown/tracker.

Cabral, E. D., & Barbosa, J. M. N. (2005). Students' opinions on the use of computer rooms for teaching anatomy. *International Journal of Morpholo, 23*(3), 267–270.

Chan, M., & Black, J. B. (2006). *Direct-manipulation animation: Incorporating the haptic channel in the learning process to support middle school students in science learning and mental model acquisition* (pp. 64–70). Mahway, NJ: Lawrence Erlbaum Associates, Inc.

Charlton, R., & Smith, G. (2000). Undergraduate medical students' views on the value of dissecting. *Medical Education, 34*(11), 961.

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences, 3*(9), 345–351.

Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A. K., … Belliveau, J. W. (1996). Changes in cortical activity during mental rotation. A mapping study using functional MRI. *Brain, 119*, 89–100.

Day, S. B., & Goldstone, R. L. (2011). Analogical transfer from a simulated physical system. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 551–567.

Dyer, G. S., & Thorndike, M. E. (2000). Quidne mortui vivos docent? The evolving purpose of human dissection in medical education. *Academic Medicine. Journal of the Association of American Medical Colleges, 75*, 969–979.

Ekstrom, R. B. R., French, J. J. W., Harman, H. H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests (Rev)*. Princeton, New Jersey: Educational Testing Service.

Finkelstein, P., & Mathers, L. (1990). Post-traumatic stress among medical students in the anatomy dissection laboratory. *Clinical Anatomy, 3*(3), 219–226.

Ganis, G., Keenan, J. P., Kosslyn, S. M., & Pascual-Leone, A. (2000). Transcranial magnetic stimulation of primary motor cortex affects mental rotation. *Cerebral Cortex* (Vol. 10,(2), 175–180. New York, N.Y.: 1991.

Garg, A. X., Norman, G. R., Eva, K. W., Spero, L., & Sharan, S. (2002). Is there any real virtue of virtual reality?: the minor role of multiple orientations in learning anatomy from computers. *Academic Medicine. Journal of the Association of American Medical Colleges, 77*(10), 97–99.

Garg, A. X., Norman, G. R., Spero, L., & Maheshwari, P. (1999). Do virtual computer models hinder anatomy learning. *Academic Medicine, 74*(10), 87–89.

Garg, A. X., Norman, G. R., & Sperotable, L. (2001). How medical students learn spatial anatomy. *Lancet, 357*.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology, 96*(3), 424–436.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*(6), 516–522.

Han, I., & Black, J. B. (2011). Incorporating haptic feedback in simulation for learning physics. *Computers & Education, 57*(4), 2281–2290.

Hariri, S., Rawn, C., Srivastava, S., Youngblood, P., & Ladd, A. (2004). Evaluation of a surgical simulator for learning clinical anatomy. *Medical Education, 38*(8), 896–902.

Hessinger, M., Holzinger, A., Leitner, D., & Wassertheurer, S. (2008). Hemodynamic models for education in physiology. *Mathematics and Computers in Simulation, 79*(4), 1039–1047.

Holzinger, A., Kickmeier-Rust, M. D., Wassertheurer, S., & Hessinger, M. (2009). Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator. *Computers and Education, 52*(2), 292–301.

Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *Journal of Computer Assisted Learning, 22*(6), 392–404.

Hunter, P., Chapman, T., Coveney, P. V., de Bono, B., Diaz, V., Fenner, J., … Viceconti, M. (2013). A vision and strategy for the virtual physiological human: 2012 update. *Interface Focus, 3*(2), 20130004.

Hu, A., Wilson, T., Ladak, H., Haase, P., Doyle, P., & Fung, K. (2010). Evaluation of a three-dimensional educational computer model of the larynx: Voicing a new direction. *Journal of Otolaryngology - Head and Neck Surgery, 39*(3), 315–322.

Jeanquartier, F., Jean-Quartier, C., Cemernek, D., & Holzinger, A. (2016). In silico modeling for tumor growth visualization. *BMC Systems Biology, 10*(1), 59.

Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P., & Montello, D. R. (2008). Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science, 32*(7), 1099–1132.

Kerfoot, B. P., Masser, B. A., & Hafler, J. P. (2005). Influence of new educational technology on problem-based learning at Harvard Medical School. *Medical Education, 39*(4), 380–387.

Kirsh, D. (1997). Interactivity and multimedia interfaces. *Instructional Science, 25*, 79–96.

Kosslyn, S. M., DiGirolamo, G. J., Thompson, W. L., & Alpert, N. M. (1998). Mental rotation of objects versus hands: Neural mechanisms revealed by positron emission tomography. *Psychophysiology, 35*(2), 151–161.

Kosslyn, S. M., Thompson, W. L., & Wraga, M. (2001). Imagining rotation by endogenous versus exogenous forces: Distinct neural mechanisms. *NeuroReport, 12*.

Lee, E. A. L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers and Education, 79*, 49–58.

Lemole, M. G., Banerjee, P. P., Luciano, C., Neckrysh, S., & Charbel, F. T. (2007). Virtual reality in neurosurgical education: Part-task ventriculostomy simulation with dynamic visual and haptic feedback. *Neurosurgery, 61*(1), 1–8.

Luursema, J.-M., Verwey, W. B., Kommers, P. a. M., Geelkerken, R. H., & Vos, H. J. (2006). Optimizing conditions for computer-assisted anatomical learning. *Interacting with Computers, 18*(5), 1123–1138.

McLachlan, J. C., Bligh, J., Bradley, P., & Searle, J. (2004). Teaching anatomy without cadavers,. *Medical Education, 38*(4), 418–424.

Meijer, F., & van den Broek, E. L. (2010). Representing 3D virtual objects: Interaction between visuo-spatial ability and type of exploration. *Vision Research, 50*(6), 630–635.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers and Education, 70*, 29–40.

Millar, H. (2016). *Can virtual reality emerge as a tool for conservation? The Guardian*. June 28. Retrieved from https://www.theguardian.com/environment/2016/jun/28/can-virtual-reality-emerge-as-a-tool-for-conservation.

Nicholson, D. T., Chalk, C., Funnell, W. R. J., & Daniel, S. J. (2006). Can virtual reality improve anatomy education? A randomised controlled study of a computer-generated three-dimensional anatomical ear model. *Medical Education, 40*(11), 1081–1087.

Parsons, L. M. (1987a). Imagined spatial transformation of one's body. *Journal of Experimental Psychology: General, 116*(2), 172–191.

Parsons, L. M. (1987b). Imagined spatial transformations of one's hands and feet. *Cognitive Psychology, 19*(2), 178–241.

Parsons, L. M., Fox, P. T., Downs, J. H., Glass, T., Hirsch, T. B., Martin, C. C., Jerabek, P. A., & Lancaster, J. L. (1995). Use of implicit motor imagery for visual shape discrimination as revealed by PET. *Nature, 375*(6526), 54–58.

Petit, L. S., Pegna, A. J., Mayer, E., & Hauert, C.-A. (2003). Representation of anatomical constraints in motor imagery: Mental rotation of a body segment. *Brain and Cognition, 51*(1), 95–101.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist, 27*(1), 91–109.

Robison, R. A., Liu, C. Y., & Apuzzo, M. L. J. (2011). Man, mind, and machine: The past and future of virtual reality simulation in neurologic surgery. *World Neurosurgery, 76*(5), 419–430.

Russell, W. M. S., & Burch, R. L. (1959). *The principles of humane experimental technique*. London: Methuen.

Schwoebel, J., Friedman, R., Duda, N., & Coslett, H. B. (2001). Pain and the body schema evidence for peripheral effects on mental representations of movement. *Brain, 124*, 2098–2104.

Seymour, N. E., Gallagher, A. G., Roman, S. A., O'Brien, M. K., Bansal, V. K., Andersen, D. K., et al. (2002). Virtual reality training improves operating room performance: Results of a randomized, double-blinded study. *Annals of Surgery, 236*(4), 458–463.

Shepard, R. N., & Cooper, L. A. (1982). *Mental Images and their transformations*. New York: Cambridge University Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology, 101*(3), 545–560.

Stull, A. T., Hegarty, M., & Mayer, R. E. (2009). Getting a handle on learning anatomy with interactive three-dimensional graphics. *Journal of Educational Psychology, 101*(4), 803–816.

Tannenbaum, J., & Bennett, B. T. (2015). Russell and Burch's 3Rs then and now: The need for clarity in definition and purpose. *Journal of the American Association of Laboratory Animal Science, 54*(2), 120–132.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599–604.

Vitale, J. M., Swart, M. I., & Black, J. B. (2014). Integrating intuitive and novel grounded concepts in a dynamic geometry learning environment. *Computers & Education, 72*, 231–248.

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition, 68*(1), 77–94.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636.

Wohlschlager, A., & Wohlschlager, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance, 24*(2), 397–412.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 22*, 87–122.