Interaction Design and the Role of Spatial Ability in Moderating Virtual Molecule Manipulation Performance

Trevor J. Barrett (trevor.barrett@psych.ucsb.edu)
Department of Psychological and Brain Sciences, UCSB
Santa Barbara, CA 93106 USA

Mary Hegarty (mary.hegarty@psych.ucsb.edu)
Department of Psychological and Brain Sciences, UCSB
Santa Barbara, CA 93106 USA

Abstract

Virtual models are increasingly employed in STEM education to foster learning about spatial phenomena. However, the role of design and spatial ability in moderating performance are not yet well understood. We examined the effects of display fidelity (stereo vs. mono), interface location (colocated vs. displaced), and spatial ability on performance during a virtual molecule manipulation task. The results indicated a significant beneficial effect of providing stereo viewing on response time, while interface location had no effect. The effect of providing stereo on performance was moderated by spatial ability. Notably, providing stereo did not benefit higher spatial ability participants, while those with lower spatial ability uniquely benefited from using the higher fidelity stereo display.

Keywords: spatial cognition; individual differences; stereo; colocation; display; interface; virtual; rotation

Computer-based virtual models are now an important instructional medium in science, technology, engineering, and mathematics (STEM) education (Trindade, Fiolhais & Almeida, 2002). Three-dimensional (3D) virtual models have shown promise in fostering meaningful learning; however, the perceptual cues provided often vary from system to system, leaving much to be understood regarding the impact of interaction design on reasoning and learning. Effects of virtual model design elements depend greatly on the given task as well as individual ability level. Given the increasing availability of new display and interaction technologies, building generalizable theories will be essential in understanding how to best design cognitively supportive virtual environments for education and research.

Here, we examine the effects of display fidelity (stereoscopic vs monoscopic viewing), interface location (collocated vs displaced), and spatial ability on a virtual object manipulation task. Regarding display fidelity, stereo displays use binocular disparity to create the illusion of depth when viewing 3D content, whereas traditional displays provide only monocular depth cues. Regarding interface location, motion tracked interfaces for manipulating virtual objects may be displaced from the virtual image (e.g. typical computer mouse position) or colocated with the virtual object to minimize disparity between visual and haptic information. In the virtual object rotation task studied in our experiment, students used a direct manipulation interface to match the orientation and configuration of 3D molecular models.

The literature suggests mixed effects of stereo viewing in virtual environments. Some studies report significant performance benefits from providing stereo viewing (Wang, MacKenzie, Summers, & Booth, 1998; Arsenault & Ware, 2004); while other studies report null effects of providing stereo (Khooshabeh & Hegarty, 2010; Barrett & Hegarty, 2013). The literature regarding colocation of haptic and visual workspaces is similarly inconsistent, with some studies reporting significant performance benefits (Ware & Rose, 1999; Barrett & Hegarty, 2013), and others reporting null effects (van Liere, Martens, Kok & van Tienen, 2005). Significant interactions between display and interface technologies have also been reported (Ragan, Kopper, Schuchardt & Bowman, 2012). The heterogeneity of findings in the virtual interaction literature likely arises from differences in experimental task demands. For example, performance benefits from providing stereo viewing are likely to be observed only when the added fidelity of the third dimension provides relevant or necessary information required for the particular task. Thus, we should be cautious in generalizing the effects of display and interface technologies, as their merit depends on the specific characteristics of the task at hand.

Virtual technologies may also be differentially effective for people of different spatial abilities. While it is widely accepted that individual differences in spatial ability play an important role in learning from 3D virtual environments, the nature of aptitude-treatment-interactions (ATI) are widely disputed in the literature (Hoffler, 2010). Two main ATI hypotheses have been offered in studies of multimedia learning, animation, and interactivity. The *ability-as-enhancer* hypothesis predicts that high ability individuals are uniquely able to utilize increased fidelity to improve performance, whereas low ability individuals do not profit due to cognitive overload. The *ability-as-compensator* hypothesis predicts that individuals with high ability are able to compensate for lower fidelity representations and do not benefit from increased fidelity, whereas individuals with low ability benefit from increased fidelity because rich external representations can supplant or compensate for lack of ability. Of course these two hypotheses are not mutually

exclusive, and may characterize different learning situations or different degrees of fidelity in the display and interface for a given task. Here, we examine the moderating role of spatial ability on the effect of display (stereo vs. mono) and interface (colocated vs. displaced) fidelity.

The task used in the present study is couched in organic chemistry, a domain that depends greatly on understanding representations of the 3D structure of molecules. Virtual molecular models are commonly employed in chemistry education (Barnea & Dori, 2000, Limniou, Roberts & Papadopoulos, 2008), making organic chemistry representations an excellent real world task domain for studying learning in virtual environments. In a previous study (Barrett & Hegarty, 2013), a more domain-specific task was employed in which participants manipulated a virtual molecular model to match the orientation of a simultaneously displayed 2D diagram representing the 3D structure of the molecule. We found a significant benefit of colocating the visual and haptic information, but no effect of stereo viewing. This task involved understanding disciplinary diagrams in addition to virtual object rotation, so it is difficult to generalize these findings to other object rotation tasks (e.g. Ruddle & Jones, 2001).

In order to improve generalizability, the task in the present study required manipulation of a virtual molecular model to match the orientation and internal configuration of a simultaneously displayed 3D model of the same molecule. This task involved two manipulations of the virtual model: 1) a global rotation of the entire molecule and 2) a local twisting rotation of a bond within the molecule. The additional local rotation made this task more complex than typical virtual object rotation tasks that involve only a global rotation of a 3D model (e.g. Ruddle & Jones, 2001). By using 3D models rather than 2D diagrams for target orientations, the present task focused on the virtual manipulation itself, and was not confounded by participants' ability to interpret diagrams.

Klatzky, Wu, Sheldon & Stetten (2008) suggest that interfaces providing perceptually mediated interaction foster better performance and learning than cognitively mediated interfaces. Perceptual mediation allows cognitively demanding processes to be offloaded onto more automatic systems; this frees up cognitive resources to be re-allocated towards performance and/or learning. Typically, when we directly perceive and manipulate physical objects, stereoscopic depth cues and congruence between vision and touch are present. Therefore, we expected that providing stereo viewing and colocation would perceptually mediate the interaction, and lead to better performance than the more cognitively mediated monoscopic display and displaced interface. In line with the ability-as-compensator hypothesis, we predicted that low spatial ability participants should differentially benefit from the increased perceptual mediation afforded by stereo viewing and colocation, while high spatial ability participants should receive little or no benefit. That is, greater spatial working-memory resources (Shah & Mikake,

1996) should allow high spatial ability participants to more effectively handle the increased cognitive demand of the mono display and displaced interface. Further, increased perceptual mediation should allow low spatial ability participants to perform more similarly to participants with high spatial ability.

Virtual Model System

The virtual model system allowed for colocated naturalistic manipulation of a virtual molecular model in a stereoscopic display. The display was mounted horizontally and faced downward towards a mirror mounted at 45°, which projected the virtual image to the viewer. When colocated, the interface was placed on a stand in the same location as the virtual image (behind mirror, see Figure 1a). When displaced, the interface was located to the left and below the mirror in the natural mouse location (38cm displacement, hands occluded).
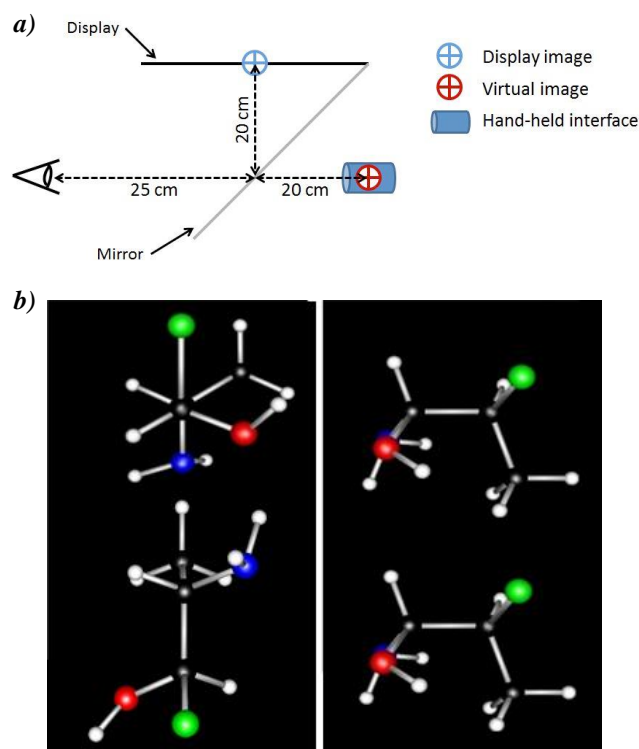

Figure 1: *a)* Mirrored display allowed for colocation of haptic and visual workspaces. *b)* Example starting orientation (left) and successful target match (right).

Nvidia 3D Vision provided stereo viewing. The hand-held interface was cylindrical and roughly the same dimensions as the virtual models. Two halves freely rotated about the long axis of the interface, allowing for local rotations within the model. A motion sensor tracked global rotations (yaw, pitch, and roll), and an optical encoder tracked local rotations. (See Stull, Barrett & Hegarty, 2012)

Method

Design

The study had a 2 (display fidelity: stereo vs. mono) X 2 (interface location: colocated vs. displaced) between subjects design, with spatial ability measured as a continuous moderating variable. Dependent measures included response time (RT) and angular error, defined as the average angular disparity of both sides of the model to the target orientation.

Participants

One hundred forty two college students (73 Female) (age: $M = 18.8$, $SD = 1.6$) participated in the study in return for course credit. None of the participants had studied organic chemistry. All had normal, or corrected to normal vision. Participants were randomly assigned to conditions.

Materials

There were 24 orientation matching trials. The starting orientation of the model was such that the global angular distance to each of the target orientations was maximized. Half of the trials involved a local rotation of the model (via rotation along the long axis of the hand-held interface). Six different virtual molecular models were used. Participants received the trials in the same order, in which two consecutive trials never showed the same molecule.

Two spatial ability measures were administered: a mental rotation test (MRT) (Vandenburg & Kuse, 1978), and a three dimensional perspective taking test, Visualization of Viewpoints (VoV) (Guay & McDaniels, 1976). Items from the NASA Task Load Index (TLX) (Hart & Staveland, 1988) assessed participants' subjective experience of the task (on a scale of 0 to 100) with regard to six criteria: mental demand, physical demand, temporal demand, own performance, effort, and frustration. Items from Waller's (2000) computer use questionnaire assessed attitudes and experience with computers. A post-experiment questionnaire collected basic demographic information.

Procedure

For the experimental task, participants manipulated the lower model to match the orientation and configuration depicted by the above target model. Participants were instructed to respond quickly and accurately; pressing a response pad key terminated the trial. The experimenter returned the interface to the starting position and participants were told to grasp (but not move) the interface before the next trial was administered.

Prior to the experimental task, participants were first given one minute to familiarize themselves with manipulating the virtual model using the hand-held interface. Participants then completed and received feedback on three practice trials that used models distinct from those in the experimental task. If unsuccessful, the experimenter demonstrated the successful

orientation match before proceeding. After the experimental task, participants then completed the NASA-TLX, MRT, VoV, computer use questionnaire, and post-experiment questionnaire.

Results

Data from 11 students were excluded from the analyses as their average error was over 2.5 standard deviations from the group mean, suggesting they did not understand the task or were unmotivated. The four interface condition groups had approximately equal numbers of males and females and did not significantly differ on the MRT, $F(3, 127) = 0.2$, $p = .90$, VoV, $F(3, 127) = 1.0$, $p = .28$, computer experience, $F(3, 127) = 0.1$, $p = .97$, or attitudes toward computers, $F(3, 127) = 0.54$, $p = .65$.

Overall, participants had an average RT of 52.8s ($SD = 19.4$). A significant effect of display fidelity (stereo vs. mono) was found on RT, $F(1, 127) = 15.6$, $p < .001$, $\eta_p^2 = .11$. Marginal means showed that participants with stereoscopic displays were about 13s faster ($M = 46.4s$, $SD = 18.8$), than those with monoscopic displays ($M = 59.1s$, $SD = 17.9$). No significant effect of interface location ($F(1, 127) = 2.2$, $p = .14$), or interaction between display fidelity and interface location ($F(1, 127) < 0.1$) was observed.
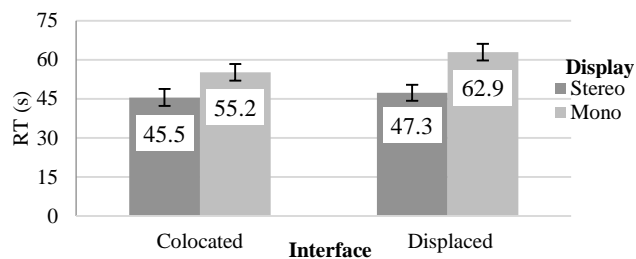

Figure 2: Effects of display fidelity and interface location on participant RT ($M \pm SE$).

Overall, participants had an average angular error of 4.8° ($SD = 1.4$). No significant effects of display fidelity ($F(1, 127) < 0.1$), interface location ($F(1, 127) < 0.1$), or interaction between display fidelity and interface location ($F(1, 127) < 0.1$) were observed.
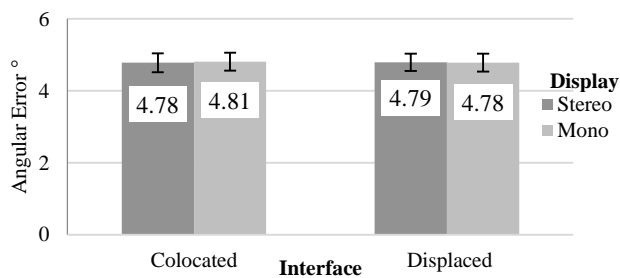

Figure 3: Effects of display fidelity and interface location on participant angular error ($M \pm SE$).

In order to account for individual differences in speed/accuracy prioritization, a composite performance score was created by standardizing the sum of standardized RT and angular error values. Note that lower values represent better performance for this measure, i.e. shorter RT and less angular error. Composite performance for the different experimental groups is shown in Figure 4. A significant effect of stereo was found on composite performance, $F(1, 127) = 8.49, p = .004$, $\eta_p^2 = .06$. Participants provided with the stereo display had better performance than those using the mono display by $0.5SD$ ($SE = 0.12$). There was no significant effect of interface location ($F(1, 127) = 1.31, p = .29$), or interaction between display fidelity and interface location on performance ($F(1, 127) = 0.37, p = .54$).
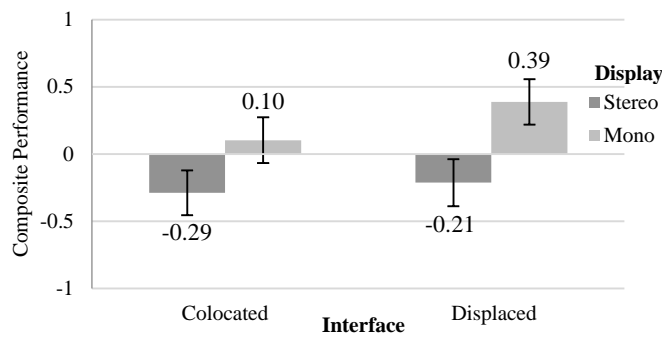

Figure 4: Effects of display fidelity and interface location on composite performance ($Z \pm SE$).

We conducted a simultaneous multiple regression analysis to examine how well the MRT and VoV scores predict composite task performance. Together, MRT and VoV explained approximately 18% of the variance in performance, $R = .424, F(2, 128) = 14.0, p < .001$. Both MRT ($r = -.345, p = .029$) and VoV ($r = -.385, p = .003$) were significantly correlated with performance, Examination of the partial regression coefficients revealed that MRT scores uniquely explained 3% of the variance in performance and VoV scores uniquely explained 6%. Scores on the MRT and VoV spatial ability measures were moderately correlated ($r = .55, p < .001$).

A composite spatial ability measure was created by standardizing the sum of standardized MRT and VoV scores. This measure was then used as a predictor in regression models that examined the interaction of spatial ability with display and interface fidelity.

To examine whether the relationship between performance and display fidelity (stereo vs. mono) and interface location (colocated vs. displaced) was moderated by spatial ability, display fidelity and spatial ability were entered simultaneously in to a regression model while controlling for interface location (entered as covariate in model). Results showed significant main effects of both spatial ability (B = -.538, SE_B = .010 $p < .001$) and display fidelity (B = -.438, SE_B = .154, $p = .005$) on performance. Participants with high

spatial ability and those provided with stereo viewing were more likely to have better performance. In addition, the interaction of spatial ability and display fidelity (B = -.345, SE_B = .16) explained 3% additional variance in performance ($F(1,126) = 4.78, p = .031$), suggesting the effect of stereo on performance was moderated by spatial ability. To explore this interaction, the conditional effects of display fidelity on performance were plotted across the observed range of spatial abilities. As shown in Figure 4, the effect of providing stereo was evident for low spatial ability individuals, while no effect of stereo was observed on those with high spatial ability. The moderator value .34 defined the Johnson-Neyman significance region, suggests a significant effect of stereo on performance for individuals below the 63rd percentile of spatial ability level. The conditional effects of display fidelity plotted across spatial ability levels are shown in Figure 5.
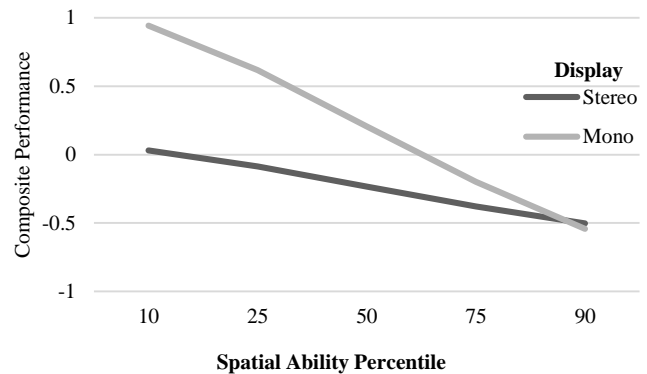

Figure 5: Conditional effects of display fidelity on performance across spatial ability levels.

In order to estimate a possible moderating effect of spatial ability on the effect of interface location we also conducted an analysis with colocation as the predictor variable (controlling for stereo), however no main effect or interaction with spatial ability was observed.

Although males scored higher on both measures of spatial ability (MRT: $F(1,129) = 6.33, p = .013, \eta_p^2 = .05$; VoV: $F(1,129) = 13.75, p < .001, \eta_p^2 = .10$), there were no significant differences between males and females on RT, $F(1,129) = 2.70, p = .10$, accuracy, $F(1,129) = 1.74, p = .19$, or composite performance, $F(1,129) = 0.58, p = .81$.

On the subjective measure of task load (NASA-TLX), participants provided with the stereo display reported significantly less temporal demand ($M = 41.2, SD = 30.9$) than those with the mono display ($M = 50.5, SD = 30.9$), $F(1,127) = 6.2, p = .01, \eta_p^2 = .04$ indicating that they were aware of the increased efficiency observed with stereoscopic displays. Participants who received the colocated interface reported significantly less frustration ($M = 32.7, SD = 40.1$) during the task than those with the displaced interface ($M = 43.3, SD = 40.1$). No significant differences were observed on the remaining items, and no interactions were observed.

Discussion

We examined the effects of varying display and interface fidelity and the role of spatial ability on performance during a virtual molecule manipulation task in which global and local rotations were required to match 3D target orientations. Response times upwards of one minute were relatively high for a virtual object rotation task, and suggested that the task was challenging to participants. Overall, providing stereo viewing lead to faster task completion. This suggests that increased display fidelity afforded by stereo viewing increased perceptual mediation of the interaction, decreased demand on spatial cognitive processing, and allowed for re-allocation of resources to benefit performance. This finding was supported by subjective reports of feeling less rushed when using the stereo display. All participants performed with a high degree of near ceiling accuracy, which was unsurprising given the lack of time constraint. Colocating the visual and haptic workspaces did not significantly benefit performance during the task. While interface location did not affect performance, participants reported feeling less frustrated when using the colocated interface.

As expected, spatial ability affected task performance; individuals with higher spatial ability performed at a higher level than those with lower spatial ability. Moreover, the predicted interaction between spatial ability and display fidelity on performance was observed, as demonstrated by a differential effect of display fidelity across spatial ability levels. Providing stereo viewing lead to increased performance for participants in approximately the bottom $2/3^{rds}$ of spatial ability, while no effect was found on participants in the top $1/3^{rd}$ of spatial ability. There was no significant interaction between spatial ability and interface location.

The more perceptually mediated stereo display benefited participants with low to average spatial ability, whereas participants with high spatial ability were able to effectively compensate for the lower fidelity afforded by the mono display. This finding was in line with the ability-as-compensator hypothesis. Higher spatial ability participants may have been able to handle the increased demand on spatial processing associated with using the cognitively mediated mono display, so that the increased perceptual mediation of the stereo display did not decrease cognitive demand to a degree that impacted performance. In contrast, lower ability participants appear to have lacked spatial processing resources required to effectively handle the increased demand associated with using the cognitively mediated mono display, while the increased perceptual mediation of the stereo display significantly decreased cognitive demand to produce increased performance.

This work adds to the body of literature on aptitude-treatment-interactions by demonstrating a clear compensating effect of spatial ability differentially moderating performance with interactions of varying fidelity. Although ATIs in virtual object manipulation are understudied, the literature regarding the nature of ATIs spans several domains. Previous studies in other domains have found ATIs supporting the ability-as-compensator hypothesis by demonstrating that low ability individuals benefited more than high ability individuals from the spatial contiguity effect (Lee, 2007), and when viewing animations rather than diagrams (Hays, 1996; Höffler & Leutner, 2011). Other studies have found support for the ability-as-enhancer hypothesis by demonstrating that high ability individual uniquely benefit when afforded additional verbal information (Mayer & Sims, 1994), visualization interactivity (Keehner, Montello, Hegarty & Cohen, 2004), and 3D models in virtual biology lessons (Huk, 2006).

The present finding that increased display fidelity differentially benefited lower spatial ability participants, combined with null effect on high spatial ability participants suggests that increasing display fidelity at least partially supplanted the moderating role of spatial ability during the task. Given that spatial ability is a relatively constant trait and the design of a virtual representation is engineered, perhaps it is useful to interpret this finding as display fidelity compensating for low spatial ability. Interpreting ability-as-compensator ATIs in this manner may offer important implications for designers of virtual interactions and learning environments who seek to improve task performance and/or learning in individuals with lower spatial ability.

Given the limitation of only finding effects on RT, future research needs to look at more complex tasks where the issue of accuracy be of greater importance. This will be especially important in order to produce stronger educationally relevant design implications. While further studies are warranted to test whether this ATI would generalize to complex tasks with learning outcomes, this finding demonstrates a clear example of interaction design differentially bolstering performance in lower spatial ability participants.

Another important point to consider is that a threshold likely exists where increasing fidelity of any aspect of the interaction (that does not increase perceptual mediation) may cognitively overload lower ability individuals, and reverse the moderation of ability on performance. Once this threshold is crossed, it is expected that ability would play the role of an enhancer in moderating performance, and high ability individuals would uniquely benefit. Investigating the nature of this threshold for different taxonomies of perceptual and information fidelity will be essential in clarifying the role of various spatial abilities in virtual interaction and learning environments. In any case, the present research highlights the importance of considering individual differences in spatial ability in the design of cognitively supportive virtual interactions.

Acknowledgements

References

Arsenault, R., & Ware, C. (2004). The importance of stereo and eye-coupled perspective for eye-hand coordination in fish tank VR. *Presence: Teleoperators & Virtual Environments*, *13*(5), 549-559.

Barrett, T. J., & Hegarty, M. (2013) Interface design and spatial cognition: A case of virtual molecule manipulation. *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 157-162). Austin, TX: Cognitive Science Society.

Barnea, N., & Dori, Y.J. (2000). Computerized molecular modeling - The new technology for enhancing model perception among chemistry educators and learners. *Chemistry Education: Research and Practice in Europe, 1*, 109-120.

Guay, R., & Mc Daniels, E. (1976). The visualization of viewpoints. *West Lafayette, IN: The Purdue Research Foundation (as modified by Lippa I., Hegarty M., & Montello DR, 2002)*.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload*, *1*, 139-183.

Hays, T. A. (1996). Spatial abilities and the effects of computer animation on short-term and long-term comprehension. *Journal of Educational Computing Research, 14*, 139–155.

Hegarty, M., & Kriz, S. (2008). Effects of knowledge and spatial ability on learning from animation. *Learning with animation: Research implications for design*, 3-29.

Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualizations—a meta-analytic review. *Educational psychology review*, *22*(3), 245-269.

Höffler, T. N., & Leutner, D. (2011). The role of spatial ability in learning from instructional animations–Evidence for an ability-as-compensator hypothesis. *Computers in human behavior*, *27*(1), 209-216.

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

Khooshabeh, P., & Hegarty, M. (2010). Inferring cross-sections: When internal visualizations are more important than properties of external visualizations. *Human–Computer Interaction*, *25*(2), 119-147.

Keehner, M., Montello, D. R., Hegarty, M., & Cohen, C. (2004). Effects of interactivity and spatial ability on the comprehension of spatial relations in a 3D computer visualization. In *annual meeting of the Cognitive Science Society, Chicago, IL*.

Klatzky, R. L., Wu, B., Shelton, D., & Stetten, G. (2008). Effectiveness of augmented-reality visualization versus cognitive mediation for learning actions in near space. *ACM Transactions on Applied Perception (TAP)*, *5*(1), 1.

Lee, H. (2007). Instructional design of web-based simulations for learners with different levels of spatial ability. Instructional Science, 35, 467–479.

Limniou, M., Roberts, D., & Papadopoulos, N. (2008). Full Immersive Virtual Environment CAVE™ in Chemistry Education. *Computers & Education, 51*, 584-593.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of educational psychology*, *86*(3), 389.

Ragan, E. D., Kopper, R., Schuchardt, P., & Bowman, D. A. (2013). Studying the Effects of Stereo, Head Tracking, and Field of Regard on a Small-Scale Spatial Judgment Task. *IEEE Transactions on Visualization and Computer Graphics*, 886-896.

Ruddle, R. A., & Jones, D. M. (2001). Movement in cluttered virtual environments. *Presence: Teleoperators and Virtual Environments*, *10*(5), 511-524.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4.

Stull, A. S., Barrett, T., Hegarty, M. (2012) Design of a virtual reality system for the study of diagram use in organic chemistry. *Proceedings of the Workshop on Technology-Enhanced Diagrams Research* (pp. 33-41)*.* Canterbury, UK.

Trindade, J., Fiolhais, C., & Almeida, L. (2002). Science learning in virtual environments: a descriptive study. *British Journal of Educational Technology*, *33*(4), 471-488.

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

van Liere, R., Kok, A., Martens, J. B., & van Tienen, M. H. A. V. (2005). Interacting with molecular structures: user performance versus system complexity. In *Eurographics Symposium on Virtual Environments, (Aalborg University, Denmark)* (pp. 147-156).

Waller, D. (2000). Individual differences in spatial learning from computer-simulated environments. *Journal of Experimental Psychology: Applied*, *6*(4), 307.

Wang, Y., MacKenzie, C. L., Summers, V. A., & Booth, K. S. (1998, April). The structure of object transportation and orientation in human-computer interaction. In *Conference on Human Factors in Computing Systems: Proceedings of the SIGCHI conference on Human factors in computing systems* (Vol. 18, No. 23, pp. 312-319).

Ware, C., & Arsenault, R. (2004). Frames of reference in virtual object rotation. In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization* (pp. 135-141). ACM.

Ware, C., & Rose, J. (1999). Rotating virtual objects with real handles. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *6*(2), 162-180.