Evaluating the Effect of Tangible Virtual Reality on Spatial Perspective Taking Ability

Jack Shen-Kuen Chang[1,2], Georgina Yeboah[1], Alison Doucette[1],
Paul Clifton[1,4], Michael Nitsche[4], Timothy Welsh[3], Ali Mazalek[1]
[1]Synaesthetic Media Lab, Ryerson University, Toronto, ON, Canada
[2]Polytechnic Institute, Purdue University, West Lafayette, IN, USA
[3]Faculty of Kinesiology & Physical Education, University of Toronto, Toronto, ON, Canada
[4]Digital Media, Georgia Institute of Technology, Atlanta, GA, USA
JackSKChang@gmail.com, gyeboah@ryerson.ca, alison.doucette@ryerson.ca,
paulgclifton@gmail.com, michael.nitsche@gatech.edu, t.welsh@utoronto.ca, mazalek@ryerson.ca

ABSTRACT

As shown in many large-scale and longitudinal studies, spatial ability is strongly associated with STEM (science, technology, engineering, and mathematics) learning and career success. At the same time, a growing volume of research connects cognitive science theories with tangible/embodied interactions (TEI) and virtual reality (VR) to offer novel means to support spatial cognition. But very few VR-TEI systems are specifically designed to support spatial ability, nor are they evaluated with respect to spatial ability. In this paper, we present the background, approach, and evaluation of TASC (Tangibles for Augmenting Spatial Cognition), a VR-TEI system built to support spatial perspective taking ability. We tested 3 conditions (tangible VR, keyboard/mouse, control; n=46). Analysis of the pre/post-test change in performance on a perspective taking test revealed that only the VR-TEI group showed statistically significant improvements. The results highlight the role of tangible VR design for enhancing spatial cognition.

CCS CONCEPTS

• **General and reference** → Experimentation; • **Human-centered computing** → Virtual reality; • **Hardware** → Tactile and hand-based interfaces

KEYWORDS

Spatial ability; spatial cognition; evaluation; tangible interaction; virtual reality; STEM education; embodied cognition; games

ACM Reference format:

Jack Shen-Kuen Chang, Georgina Yeboah, Alison Doucette, Paul Clifton, Michael Nitsche, Timothy Welsh, and Ali Mazalek. 2017. Evaluating the Effect of Tangible Virtual Reality on Spatial Perspective Taking Ability. In *Proceedings of the 2017 Symposium on Spatial User Interaction, Brighton UK, October 2017 (SUI '17),* 10 pages.
https://doi.org/10.1145/3131277.3132171

SUI '17, October 16–17, 2017, Brighton, United Kingdom

ACM ISBN 978-1-4503-5486-8/17/10…$15.00
https://doi.org/10.1145/3131277.3132171

1 INTRODUCTION

In many large-scale and longitudinal studies, spatial ability has been shown to be a strong predictor for STEM learning and career success. For example, SMPY (Study of Mathematically Precocious Youth) tracked 5,000 students who performed well in school for 35 years starting from 1971, and showed the strong association between spatial ability and success in STEM fields [25,33]. Wai et al. [37], in 2009, examined 50 years of similar research, with samples obtained from 400,000 participants who were followed up in the studies for at least 11 years. Wai's findings were coherent with those from SMPY's, and reinforced the notion that spatial ability can be a strong predictor of STEM performance.

However, many existing spatial ability testing and training materials, albeit well-tested by cognitive scientists and widely-used by educators, do have certain limitations: 1) They are presented on a surface (e.g., paper or computer monitor), which does not best afford 3D perception; 2) They do not encourage the use of visuo-motor skills during the spatial problem-solving process, e.g., a student may only sit still with very little bodily movements while using pens, keyboards, and mice; 3) The content or visuals in those materials are often not appealing to students (e.g., not immersive and engaging). These limitations show that using those materials can be somewhat different from how people actually use spatial ability to interact with real-world objects and environments.

The concept of "embodiment" has become important in cognitive science since the mid-1980's [40]. It is based on the principle that "cognitive processes are deeply rooted in the body's interactions with the world" [39]. Given recent research that has shown a link between embodiment and spatial cognition, there is much potential to address the aforementioned limitations by studying how an interactive system that leverages embodied and tangible interaction can engage, and even improve spatial ability.

TASC (Tangibles for Augmenting Spatial Cognition) is a system that establishes embodiment in several ways: head-tracking, hand-tracking, and tactile input with tangible blocks [2]. It supports embodied interaction as it tasks users to solves spatial puzzles in a specially designed VR environment that packages those interactions. We described TASC's design process, particularly how it was designed to support the use of spatial perspective taking ability. Besides the iterative/formative evaluations during the design process, the study also evaluated the final design's interface feasibility. That summative evaluation showed a positive experience among participants and noted ways in which they were using their bodies to facilitate spatial problem solving with perspective taking ability.

In this paper, using a comparative experiment with 3 conditions, we show that TASC can improve the perspective taking ability among participating college students from STEM majors. We report on the design and experiment to study TASC's embodiment effect on perspective taking abilities. We then present an analysis of the quantitative data. Finally, we discuss design implications and projected effects on future possibilities. This exploratory study makes contributions to the fields of spatial cognition as well as interaction design in its exploration of the effects that tangible VR interfaces provide for spatial cognition.

2 RELATED WORK

Spatial cognition through embodied cognition and common coding has been shown as an active form of engagement across different scales, within different forms of embodiment, and in different interaction settings. However, how to specifically design for and assess these effects remains a challenge, or at least, an area not well-explored yet.

2.1 Embodied Cognition & Common Coding

Embodied cognition has been studied in many fields such as ontology, phenomenology, social interactions, etc. While this research originates from and focuses on the philosophical aspects [27], we pay more attention to the cognitive science side of embodied cognition. In general, embodied cognition provides a holistic view to study how actions are correlated with the user's own intention and perception [39,40]. That is, the visuo-motor system is always part of cognitive processing, not merely an input or output of the brain. Robbins and Aydede [32] referred to embodied cognition as "embeddedness of the brain in the body", which outlined the interplay of the mind and the body's movements.

Common Coding Theory (Ideomotor Theory), a topic in embodied cognition, describes the linkage between action, perception, and cognition [14]. Furthermore, this linkage can be activated or strengthened both ways, i.e. not only can cognition and perception lead to actions, the use of motor system can also activate imagination or perception of motion. Examples: Viewers might move their arms or legs to displace the body weight distribution while watching a scene with a car making a drastic turn in movies or video games [30]; video game players move their bodies trying to dodge a bullet displayed on the screen; people can identify their own bodily movements, even if they are presented in a visually abstracted form (e.g., something as simple as a light-point animation) [20,21] .

2.2 Spatial Cognition

Spatial cognition is about how humans acquire, organize, and use spatial information in relation to objects or environments. Per Eliot [6], spatial functioning is "pervasive", i.e., it is a cognitive activity that is required almost everywhere and all the time. It affects a range of research fields but here, we select several topics related to the use of the body.

There are several categorizations to describe the use of spatial cognition in relation to the body. To describe different spatial scales with respect to the body, cognitive scientists devised the classification of figural, vista, and environment [11]. This classification is based on the interactor's body and how he perceives external objects in relation to it: figural scale – graspable, within arm's reach range; vista scale – observable area; environment scale – transvers-able territories. Specifically in the case of perspective taking, spatial ability has egocentric or allocentric types [24] (in the first type the viewer imagines perspectives from his own body as the reference point, in the second type the viewer imagines points of views from objects that are not his own body).

In addition, solving spatial tasks can be categorized as epistemic or pragmatic. The former regards actions that are not made directly for the goal, but are trials and errors to reduce work complexity (e.g., memory workload). The latter regards actions that are directly performed to achieve the task at hand. The idea of epistemic actions was introduced by Kirsh et al. [19], who used the video game Tetris to demonstrate the concept. Kirsh observed that performant Tetris players keep rotating descending pieces (exploratorily or even mindlessly) until they see a matching position or angle for the piece to drop (i.e., epistemic action), while novice or lower-skilled players often just mentally rotate a piece when seeking a matching spot for it to land. The use of the body and epistemic actions have been studied as spatial problem-solving strategies in tangible user interaction [1,7].

2.3 TEI + VR for Spatial Ability

TEI (tangible & embodied interactions) has broadly incorporated spatial manipulation and bodily movements to design post-WIMP (windows, icon, menu, and pointer) interfaces. This can be demonstrated from seminal frameworks such as the RBI frame work (Reality Based Interaction) by Jacob et al. [16], or Hornecker and Burr's [15] tangible design framework. They both unify existing research themes and point out future design opportunities.

VR (virtual reality) provides immersive environments and affords better 3D perception of the digital content. It has been used for virtual laboratories [23], education [36], fire-fighting training [38], and healthcare [22] among other cases.

It is natural to envision the potential of VR-TEI systems to support spatial ability, like the opportunities pointed out in our design framework [3]. However, currently, we find very few VR-TEI systems designed around supporting spatial ability, let alone around the necessary evaluation with respect to spatial cognition. For example, Dünser et al. showed VR and AR

(Augmented Reality) can be a good spatial ability training tool [5], but the project did not employ tangible objects. Conversely, the BDC project [26] showed that spatial mental rotation ability can be improved from operating a tangible and embodied exoskeleton, but the system did not involve VR – the users were performing virtual spatial tasks displayed on a wall surface.

Based on these systems found separately in the VR and tangible interface fields, and supported by theories from embodied and spatial cognition, we believe there is relevant potential for a combined approach to better understand the effects of a VR-TEI system's possible enhancement on spatial ability.

3 THE TASC SYSTEM

The detailed design rationales and system description of TASC are included in our previous work. Here we provide a summarized description from [2] on how TASC works, which will serve as a foundational understanding for how we designed and conducted the evaluation of the set up.

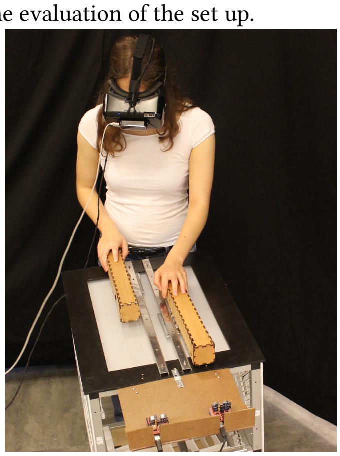

Figure 1: The physical setup of the TASC system

3.1 Target Spatial Ability: Perspective Taking

TASC was designed around one particular spatial cognitive activity: perspective taking. Perspective taking, also known as spatial orientation ability, is defined as "the ability to mentally represent a viewpoint different from one's own" [8]. This spatial ability is important to map-reading and navigation, and also a necessary skill for fields such as pilot training, engineering (e.g., drafting), architecture (e.g., 3D modeling), chemistry/physics/biology (e.g., molecule structures). Perspective taking is a good fit for evaluating a tangible VR interface's effect on spatial cognition as: 1) Perspective taking ability has been shown to be associated with bodily movements [12,34], and such movements form the basic interaction condition of tangible interfaces; 2) Many perspective taking ability tests are available [8–10,13], thus allowing for strong evaluation in context with other research; 3) Perspective taking as cognitive ability remains malleable, i.e., it can change or improve beyond a certain age instead of remaining fixed after childhood [4,28,35]. This suggests that an intervention may possibly lead to changes in this ability.

3.2 Tangible & Virtual Interactions for Embodiment

The TASC system realizes embodiment in three ways: 1) The user wears an Oculus Rift head mounted display (HMD), which tracks his head movements and immerses him in a 3D virtual environment (VE); 2) A Leap Motion sensor is attached to the front of the HMD (making them a "HMD bundle"). The Leap Motions tracks the user's hand movements and renders the captured hands in the VE; 3) Two wooden blocks placed on a table are the main controller for the virtual gameplay. The two physical blocks mimic the behavior of virtual objects (two fences) and can only move linearly, as each of them is constrained by a rail.

The system was built in Unity 5.3 (programmed with C#). To track the user's head and hand movements, Oculus Rift and Leap Motion SDKs were integrated in Unity. On the physical computing side, two ultrasonic sensors connect via an Arduino to the main computer and the Unity environment to detect the blocks' linear movements.

3.3 Gameplay

The VE includes a game that asks the user to solve a series of spatial puzzles by making use of their perspective taking ability. The VE consists of a farm (with appropriate structures such as a cabin, windmills, bushes, and a stack of logs). It includes a horse whose initial position is always separated from the user's ground character position by two long fences. The horse cannot cross over to the player as long as it us separated by the two fences. But each fence contains an opening. The goal of each game level is to move the physical blocks, so the virtual fences' openings are aligned, revealing a pathway for the horse to run toward the user's ground position. The user has two main perspectives to solve these puzzles.

The Ground View (GV) (Fig 2): This is a first-person view in which the user's virtual character is situated on the ground. In this view, the user can only look around, and cannot move around in the VE. With this view, he can see the opening of the near fence (the fence closer to his GV's position) by looking around, as well as the approximate position of the opening of the far fence (the fence closer to the horse). He can also look around to examine the surrounding objects, e.g., his spatial relationships to the cabin, the pile of log, the windmill, etc. However, the user cannot move the fences in this view. To move the fences, the user needs to switch to the Aerial View.

The Aerial View (AV) (Fig 3): This is also a first-person view, also a view within which the user can only look around and cannot move around in the VE. In this view, the user looks down onto the farm from a bird's eye view. In this view, the user receives an overview or outlook of the spatial relationship of the farm's objects: the farm's structures, the fences, the horse, and the ground character's position (where GV is located), which is a short orange cylinder with text label "You". However, the positions of the fences' openings are hidden in this view. Seen from this bird's eye view, each fence appears to be a long and continuous structure with its opening hidden from the user.

Although fence openings are hidden in AV, this view is the “action view”. Only in this view the user can change the positions of the fences by moving the physical blocks along their rails. Each block controls a corresponding fence.

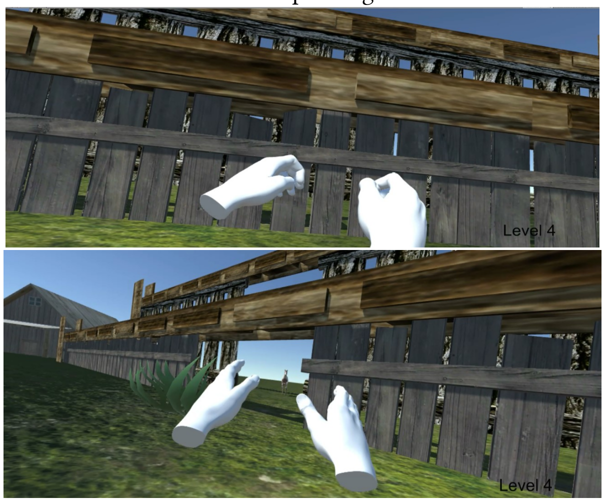

Figure 2: Ground View (GV) (Top: before the puzzle is solved; Bottom: the puzzle is solved, the horse runs toward the user.)

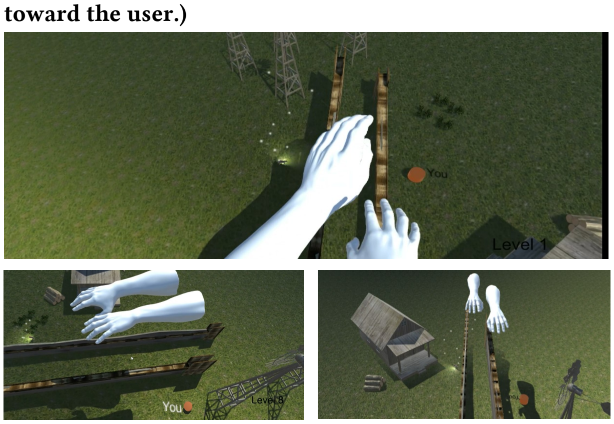

Figure 3: Aerial Views (AVs) (Top: normal, 0° view; Lower left: 90° view; Lower right: 180° view – the mirror view)

In GV, the user can see where the fences’ openings are, but cannot move the fences; in AV, the user gets to move the fences, but does not see the current positions of the openings. Both viewpoints feature a clearly marked “perspective switch” icon. The user can look at it for 0.5 seconds to switch to the other view. The game challenges the user to keep switching perspectives (GV: “solution progress view”; AV: “action view”) so he can carry the spatial information acquired in one view to the other in an iterative manner. This should eventually leads to solving the spatial puzzle by aligning the fence openings to form an open path between the horse and the user’s ground position. Only then can the horse run through the opening and towards the user character as an indication that this puzzle is solved.

There are 9 levels (9 puzzles) in total, with increasing spatial difficulty as the user progresses. Increasing the difficulty engages the user so he will not feel bored from repetitive content and interaction. The spatial difficulty is a result of a mix-and-match method of certain spatial features, e.g., the horse is placed diagonally relative to the user’s GV position (the you-icon position seen in AV); multiple and randomizing AVs (90° angle, 180° angle, plus the normal AV); hiding the you-icon to force the user to use the surrounding objects in the VE as landmarks when he is in GV. These features are aimed to continuously engage the user to keep applying his perspective taking ability, while employing more motor skills in the spatial problem-solving process.

4 EXPERIMENT OVERVIEW

In order to evaluate how (or if at all) a tangible VR system, as an intervention, can improve perspective taking ability, we designed and built an experimental setup that compared the TASC condition with 2 comparative conditions. In Section 5, we describe how certain design choices were made for building these comparative conditions. Section 6 details how we chose and digitized the pre/post-tests from an existing perspective taking test in cognitive science. In Section 7, we describe the workflow of the experiment.

5 COMPARATIVE CONDITIONS

5.1 The “Low-Embodiment” Condition

The main goal for this condition is to provide the same game content to play with but provide a lower level of embodiment. To conduct a meaningful and practical exploratory study and avoid over-complicated experimental design, we followed related research that tests how a novel “interface package” compares to a conventional Graphical User Interface (GUI) one, e.g., tangibles plus augmented reality vs. GUI [31], TUI (tangible user interface) vs. GUI [17,18], or TUI vs. a multi-touch display [1].

The full TASC version features many interaction factors that support immersion, such as tangible blocks, Oculus Rift, hand tracking with Leap Motion. To contrast this design, the “low-embodiment” condition uses a keyboard, a mouse, and a flat monitor (27'' display). This set up allows for principal interaction with the same game components but the interaction design detaches players from the objects (e.g. the moving blocks, the usage of the HMD bundle) and thus provides a low-embodiment version as a means to compare the spatial ability effects of desktop computer interactions (common in classroom settings) versus a tangible VR system. The design range between the tangible VR and the desktop setup offers a number of possible in-between conditions, which we did not deploy. Instead, the low-embodiment setup presents an ecological comparative condition (an interface package). It also mirrors the current conditions of educational facilities, that usually do not have access to tangible interfaces but rely heavily on either tablets or basic PCs with a keyboard and mouse set up in their computer labs.

In the low-embodiment condition, two pairs of keys were mapped to control the fences (one key to move one fence up, one to move it down): Q/A for one fence, E/D for the other. In both viewpoints, the participant could move the mouse to look around, similar to using a first-person controller in video games, but we excluded spatial movement to mimic the TASC setup (i.e., in both GV and AV, the user could only look around and cannot walk around). We substituted keyboard and mouse for tangible blocks and Oculus Rift. Consequently, this condition put much less emphasis on embodiment, even though it had the exact same controller output and the same game content. For example, the increasing difficulty levels were kept intact in the game design.

We found that it took roughly the same amount of time to complete the 9 levels in the low-embodiment condition as it took in the full TASC condition (around 15 to 20 minutes). Hence, we could reasonable assume that these two conditions' main influence on the participants is the interface difference, not other things such as time needed to complete.

5.2 The Control Condition

This condition did not use any of TASC's interface and game content to test for a basic learning effect between pre- and post-tests that could have emerged independently from any intervention. Thus, the control condition avoided spatial cognitive engagement. It consisted of a list of questions we presented on a Google Form shown on the same 27" screen as the low-embodiment condition's. These questions were simple math (e.g., "solve 3x + 2 = x + 4"), retyping a long word or sentence (e.g., "quintessential"), and multiple-choice questions with basic grammar (e.g., "It's okay. You did your/you're/you best and I'm very proud of you."). We designed this form so it would take around 15 to 20 minutes to complete. This duration was established to be similar to the range of interaction time in the TEI and Keyboard conditions. While this condition's participants were also mentally engaged for around 15 to 20 minutes, the questions were designed so that they did not have a spatial or visual component like a geometry problem would have. Thus, this control group enabled us to determine if there might be any pre-to-post improvement simply as a function of time and/or learning from taking the test twice. As such, we can more effectively attribute any impact of the 2 intervention groups to the characteristics of that particular set up.

6 THE PRE/POST-TESTS

We implemented the pre/post-tests for the spatial perspective finding task as a digital version of the paper-based Perspective Taking/Spatial Orientation Test (PTSOT) developed by Hegarty et al [10]. Hegarty's PTSOT was chosen because of its similarity to TASC's game mechanics: one needs to use a top-down view to re-orient one's perspectives in relation to several surrounding objects in order to find the direction of a target object with respect to the origin object and the reference object. In addition, PTOST is 2D-based (unlike others such as Frick et al.'s which has 3D visual cues [8]). This choice allows us to better evaluate the effect of an embodied and tangible experience transferred to surface-based content. Finally, certain perspective taking tests are designed for young children only, e.g., Piaget's Three Mountain Task [29], Frick et al.'s Perspective Taking Task [8], or IPT Items (Imaginary Perspective Taking) [13], but PTSOT can be administered to adults - in our case: college students.

The content of our 12-question pre-test was the same as Hegarty's PTSOT. The post-test had the same questions as the pre-test, but the questions' order was shuffled. This was to reduce familiarity effect. (In case "shuffling" causes confusion: every participant received the same pre-test, and the same post-test, i.e., a participant did not get an uniquely shuffled post-test.)

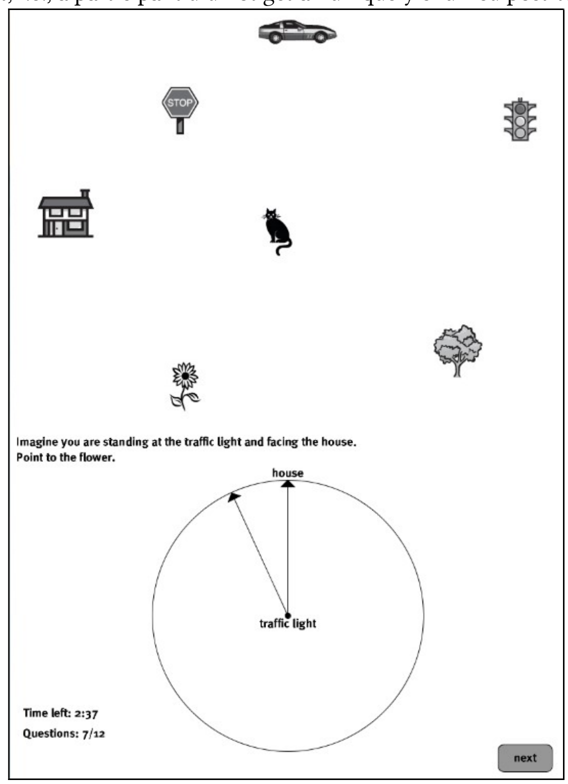

Figure 4: Our digital version of Hegarty's PTOST

We built the pre- and post-tests as desktop applications in Processing 3.0. For each question, the user drags/drops the mouse cursor to provide his answer, indicating what he thought was the spatial relationship between origin, reference, and target objects. Our application recorded and analyzed the user inputs (user solution angle, correct solution angle, completion time, etc.) much more easily than the original paper version. But overall, we adhered to the instructions of PTSOT: each participant had 5 minutes to work on the 12 questions as a full (pre- or post-) test; since our tests were administered on a fixed computer monitor, it inherently also followed PTSOT's instruction ("Please do not pick up or turn the test booklet") to keep the reference image in place in order to avoid visual aids that might affect the test.

Fig 4. shows the interface of our pre/post-test. The question asks: "Imaging you are standing at the traffic light and facing the house, point to the flower". In the answer circle, the arrow between the traffic light (Origin) and the house (Refence) is provided by the interface (same as PTSOT). Fig 4. also shows a user provides his answer by drawing another arrow from Origin to Target (flower).

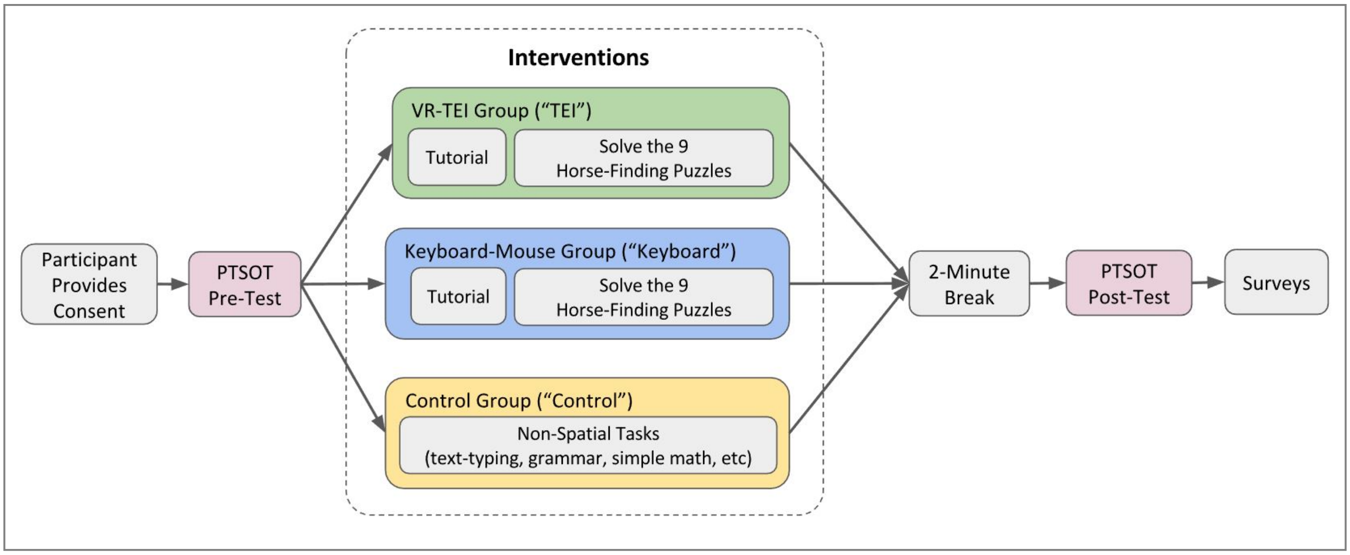

Figure 5: The experiment workflow

7 EXPERIMENT PROCEDURE

The protocol for this comparative experiment was approved by Ryerson University's ethic review board. The overall workflow for the experiment is illustrated in Figure 5. There were 3 conditions of interventions: full TASC, low-embodiment TASC with keyboard and mouse, or control (respectively TEI, Keyboard, and Control for short).

7.1 Before the Intervention

A participant, after being greeted and briefed about the study, gave consent by signing the consent form. He then was given 5 minutes to take a PTSOT pre-test. After the pre-test, he was randomly assigned to one of the 3 outlined conditions.

7.2 The Intervention

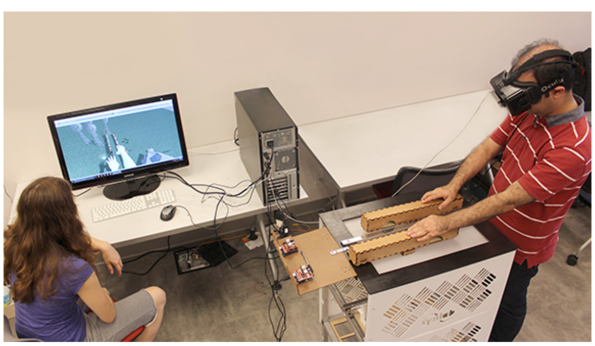

Figure 6: TEI group intervention setup (mirrored view in a later level)

(This part is similar to what we described in [2] about evaluating the participant's system interaction.) The participant was taken to the front of the intervention setup that was randomly assigned to him. In the TEI condition, the researchers helped him put on the HMD bundle, made sure it was stable and comfortable, and gave him a short tutorial. This tutorial's purpose was to establish proficiency with the novel interaction set up provided. The tutorial assured that each participant was familiar enough to use and control the TEI condition's system, and to balance out unfamiliarity with the system itself through a proficiency test phase. Establishing sufficient proficiency is a common experiment step in cognition related studies. Proficiency was determined if a participant passed the first 2 levels without major difficulties. During the tutorial, he could ask more questions to get clarifications if needed. Having passed the tutorial, he entered the main interaction session consisting of 9 levels of the horse finding game (i.e., in total he would play the first 2 levels twice).

There was no time limit for any single level but whenever a puzzle was solved by the participant, a researcher gave a verbal prompt ("Now, I am going to advance you to the next level.") and pressed a keyboard shortcut of the system's computer to advance him to the next puzzle.

The second group (Keyboard) interacted with the low-embodiment version of the system described above. A Keyboard group's participant had similar intervention workflow as TEI's: first establishing proficiency with the tutorial, and then conducting the main test through the various game levels.

The third group was the Control condition outlined above. Here, participants did not engage with any spatial task during the intervention. The Control condition was designed to test for possible learning effects in-between the pre- and post-tests.

7.3 After the Intervention

Following the intervention, each participant was provided with a 2-minute break. The reason for the break was that he could rejuvenate from any possible fatigue incurred from the intervention, while keeping the effect of the intervention fresh. The duration was chosen from our pilot study's experience, and based on our cognitive scientist's expertise. A PTSOT post-test was administered following the break. Then, the participant was asked to complete a survey about his background and demographics information.

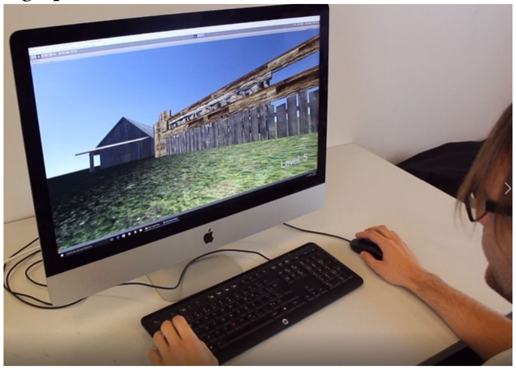

Figure 7: Keyboard (& mouse) group intervention setup (looking around with mouse in Ground View)

8 RESULTS

8.1 Overview

In total, the study involved 52 participants (6+46). For the pilot study, we had 6 participants (3M/3F) whose responses and feedback helped with the finalization of the system design (particularly the TEI and Keyboard conditions), and the experiment's protocol. For the main (formal) study, we had 46 complete participants: 15 in the TEI group (7M/8F), 16 in the Keyboard group (9M/7F), and 15 in the Control group (8M/7F). Overall, we obtained gender balance in each group. The results presented below are only from the main study's 46 participants.

Most participants were undergraduate students or recent college graduates from STEM majors, and their ages concentrate in the range of 18 to 28, with 3 participants in their 30's. Each participant in the 3 groups took around 15 to 20 minutes to complete his intervention. The results are described and discussed in the following sections with the aim to answer this question: does TASC improve participants' perspective taking spatial ability?

8.2 Statistical Analysis

The following section outlines the results of a series of analyses that were conducted to test the specific a priori predictions that completing the TEI task would improve perspective taking spatial ability (i.e., a pre/post task decrease in error), whereas performance of the Keyboard and Control protocols would not improve perspective taking spatial ability (i.e., no pre/post task decrease in error).

8.2.1 Data Reduction. To determine if completing any of the protocols altered perspective taking, each participant in the different groups completed the pre-test and post-test version of PTSOT. The answer to each question was recorded and we analyzed the *precision* and *accuracy* of their performance in each test as a measure of their perspective taking ability. The accuracy (mean difference between correct solution angle and reported angle) and precision (variability of the differences between correct solution angle and reported angle) of each participant's performance was calculated using the following procedure. First, the difference between the target angle (correct solution) and the participants' reported judgment of the angle on each trial was calculated as angular error. For *accuracy*, the mean of the absolute values of each difference score (correct angle – judged angle) was then calculated for each participant. The absolute value of the difference score was used because errors on different sides of the actual location should be treated equally (and may cancel each other out if signed values are averaged, providing an underestimation of the judgment error). For *precision*, the standard deviation of the signed difference scores was calculated for each participant. The signed difference scores were used for the calculation of precision (variability) because it provides a better assessment of the dispersion of the judgments around the target.

Prior to calculating the accuracy and precision measures, trials on which the errors were greater than 90° were removed as outliers and recording errors. These trials were considered to be outlier or interpretation errors because a difference between the actual and judged angle of greater than ±90° means that the participant placed the cursor/marker on the opposite side of the circle from the actual correct solution. Hence, these responses were likely to be a rare instance in which either there was an error in interpreting the question by the participant or simply a lack of attention. Consistent with the notion that these trials were outliers being rare trials that did not represent the participants typical performance, only 42 trials were removed from the overall data set, with an average removal of 1.7/24 (7.1%) trials per participant (24 trials: 12 questions x 2 tests).

It should be noted that, although the statistical analyses reported here used the accuracy and precision measures calculated after these outliers were removed from the data set, a secondary set of statistical analyses were performed on these measures that included these outlier data points, i.e., the whole raw dataset, including the data with ±(90 to 180)°. The results of the secondary analyses on the entire data set were consistent with the result of the analyses with the outliers removed. Although both sets of analyses are similar, the result of the analyses of the data without the outliers are reported here because we believe the analyses without the outliers are more reliable and valid, and more accurately reflect the true performance of the participants.

8.2.2 Main Statistical Analysis. To address the specific a priori predictions regarding the (non-) influence of the intervention on perspective taking, a series of pre-planned paired sample t-tests were completed. Separate t-tests were conducted to compare the accuracy and precision scores on the pre- and post-tests. These tests were conducted separately for each group. Because 6 t-tests were conducted in total, α was corrected to 0.0083 according to the conventional Bonferroni correction procedure.

To test the prediction that completion of the TEI group of TASC would improve perspective taking ability, the accuracy and precision scores from the pre- and post-TASC test were submitted to separate paired sample t-tests. The analysis of accuracy revealed a significant pre/post-intervention increase in accuracy, t(14)=3.97, p<0.0083, mean improvement 5.96°, 95% CI of difference scores = 2.74-9.18 (see Figure 8). The effect size for the TEI group's change in performance was considered to be large (d=1.02) with 13 of the 15 (87%) participants in the TEI group improving their accuracy (see Figure 9). Likewise, the analysis of the precision for the TEI group showed a significant improvement, t(14)=4.25, p<0.0083, mean improvement 8.17°, 95% CI of difference scores = 4.04-12.30, with a large effect size, (d=1.09) (see Figure 8). As can be seen in Figure 9, 13 of the 15 (87%) participants in the TEI group improved precision. Overall, these data indicate that the group that completed the TEI intervention showed a significant improvement with large effect size – a pattern of effects consistent with the predictions.

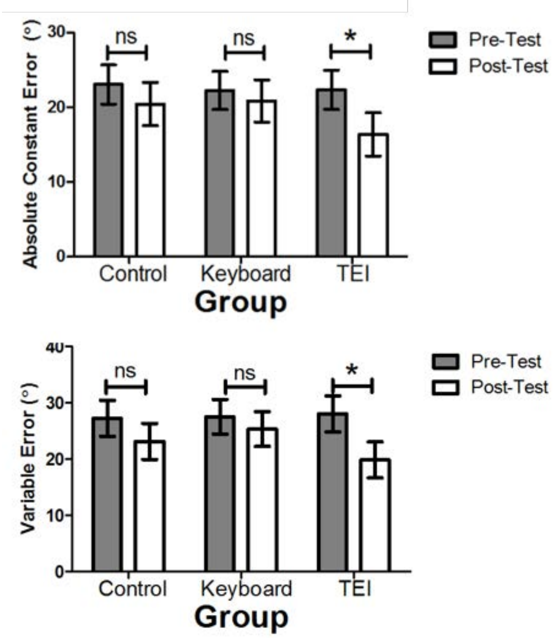

Figure 8: Mean accuracy (top panel) and precision (bottom panel) in the different groups on the pre- and post-test. (SEM bars are shown.)

To test the remaining predictions that the participants completing the Keyboard and Control interventions would not show a per/post change in their perspective taking abilities, the same set of analyses were conducted on the data from these groups. The results of the analyses of the accuracy and precision scores for the Keyboard and the Control were consistent with predictions. Specifically, there were no significant improvements in the accuracy of the Keyboard group, t(15)=0.55, p<0.59, mean improvement 1.41°, 95% CI of difference scores = -4.05-6.88, d=0.14, or the Control group, t(14)=1.04, p<0.31, mean improvement 2.61°, 95% CI of difference scores = -2.76-7.98, d=0.27. Only 10 of the 16 (62.5%) participants in the Keyboard group and 8 of 15 (53.3%) participants in the Control group showed a pre/post increase in accuracy (see Figure 9). Likewise, there were no significant improvements in the precision of the Keyboard group, t(15)=0.75, p<0.47, mean improvement 2.18°, 95% CI of difference scores = -4.03-8.48, d=0.19, or the Control group, t(14)=1.53, p<0.15, mean improvement 4.10°, 95% CI of difference scores = -1.66-9.86, d=0.39. Only 9 of the 16 (56.3%) participants in the Keyboard group and 9 of 15 (60%) participants in the Control group showed a pre/post increase in precision (see Figure 9).

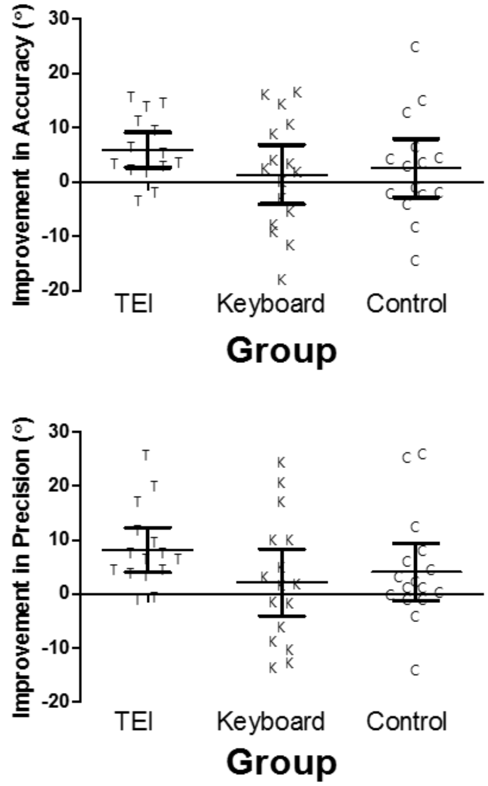

Figure 9: Plots of each individual participants' pre/post-test difference scores along with the group mean difference scores and the 95% confidence intervals for accuracy (top panel) and precision (bottom panel).

Note that, although a Bonferroni-corrected α was used to control for potential increases in Type I error as the result of multiple statistical tests that were conducted, the p values of these analyses did not even approach a more liberal α of 0.05. Overall, the results of these analyses revealed that, although there may have been some negligible performance improvements that occurred simply from performing the perspective taking task twice (an effect of test practice), these effects were in the small or medium range and, more importantly, not statistically significant. These results stand in contrast to those of the TEI group suggesting that there was a significant benefit (with a larger effect size) to perspective taking.

8.2.3 Secondary Statistical Analysis. Although we chose the above planned-comparisons approach to address the specific a priori predictions we had formed, we also conducted a series of secondary analyses to further explore the data. In one analysis, we submitted the accuracy and precision data to separate 3

(Group: TEI, Keyboard, Control) by 2 (Time: pre, post) mixed ANOVAs with Group as a between-subjects factor and Time as a within-subjects factor. These analyses would allow us to determine if there were any overall effect of time or any group differences in performance. The analysis of accuracy revealed a main effect for Time, $F(2, 43)=6.50$, $p<0.05$, that indicated that, overall, there was an improvement in accuracy from pre-test (M=22.6.0; SEM=1.50) to post-test (M=19.2; SEM=1.66). The main effect for Group, $F(2, 43)=0.28$, $p>0.7$, and the interaction between Group and Time, $F(2, 43)=1.09$, $p<0.35$, were not significant. The analysis of precision revealed a similar pattern of findings. The main effect for Time, $F(2, 43)=10.60$, $p<0.005$, indicated that, overall, there was an improvement in precision from the pre-test (M=27.6; SEM=1.82) to the post-test (M=22.8; SEM=1.83). The main effect for Group, $F(2, 43)=0.19$, $p<0.83$, and the interaction between Group and Time, $F(2, 43)=1.43$, $p<0.26$, were not significant.

Another additional analysis consisted of two sets of ANCOVAs in which the pre-test scores were used as the covariate and Group was the sole independent between-subjects factor. These analyses were conducted because it could have been that the baseline (pre-test) scores could have affected or determined the pattern of performance in the post-test. In each case, pre-test score was a significant covariate ($p<0.01$), but using it as a covariate did not change the pattern of results. That is, there was no Group difference in post-test accuracy, $F(2, 42)=1.19$, $p<0.32$, or precision, $F(2, 42)=1.56$, $p<0.23$. Thus, pre-test performance did not influence the overall pattern of effects.

Although these analyses did not reveal group differences in the performance or the change in performance (i.e., no Group main effects or Time by Group interactions), we hold that the main analyses in which the a priori predictions were directly tested were more valid and relevant. We hold that the main analysis is more relevant because any Group differences and, in particular, the Group by Time interactions will only inform whether or not the pre/post differences across the groups are significantly different (i.e., a change in one group's performance is larger than the change in another group's performance), not that there was a significant pre/post change or not in a given condition. Only separate planned comparisons of the pre/post data within each Group can provide the direct answer as to whether or a specific intervention affected performance. Further, we note that the effect size calculations for pre/post changes (Cohen's d) revealed large effects for the group experiencing the TEI training, whereas the Keyboard and Control groups had only small or medium sized changes in performance. Thus, the TEI group alone experienced a statistically significant change in performance with large effect sizes.

9 DISCUSSION

With multiple methods of analysis, we have demonstrated that TASC is indeed a system that can improve perspective taking spatial ability. However, the role of embodiment (established with tangible objects and movement-tracking virtual interactions)0 with respect to spatial ability is not very clear so far. Does embodiment proactively augment spatial ability? Or does it passively offload working (spatial) memory for spatial tasks? Which of the interaction factors in the full TASC is more influential to perspective taking ability? One way to answer these questions is to modify the existing TASC system to allow for more differentiated levels of tangibility and virtual interactions. Also, since we have obtained positive result in this initial exploratory study, it is not unreasonable to start thinking about visionary topics such as how tangible VR interactions can be integrated to STEM curriculum, Our previous work showed TASC received positive response [2]. Also, for spatial problem solving using perspective taking ability, TASC encouraged movements, verbalization, and individual development of solution strategy. In this paper, we have shown that after using full TASC, the target spatial ability improved (more statistically meaningful than in other conditions). How do we transform this short-term improvement into a long-term enhancement of spatial ability? What are the generalizable technological and design lessons we can apply to build other tangible VR systems for other spatial abilities? Last but least, how do we include domain-specific content in those systems so acquiring and using the spatial skills can be a more natural and rewarding experience? These questions drive our future work.

10 CONCLUSION

Based on theories in embodied and spatial cognition, there is potential to use tangible VR systems to engage and improve spatial ability, a cognitive ability that is highly linked to STEM learning and career performances. Very few tangible VR systems are designed to support a target spatial ability, let alone the relevant evaluation about their spatial effects on users. In this paper, we describe an experiment designed to study TASC, a tangible VR system we built for the spatial ability of perspective taking. The experiment, based on cognitive science considerations, consisted of 3 conditions: TASC (TEI), low-embodiment TASC with keyboard & mouse (Keyboard), and a control group (Control) who performed non-spatial tasks (with keyboard & mouse). Before and after a participant's assigned intervention, he took perspective taking pre- and post-tests. In total, we include the results from 46 participants, with around 15 per group. Our experiment results, examined with multiple layers of analysis, revealed that although both TEI and Keyboard groups showed improvement, only the TEI group showed statistically significant improvements in their performance (both in precision and accuracy) on the perspective-taking test following the intervention. This current result suggest that a more embodied interface may lead to better improvement in perspective taking ability. We position the research as an initial exploratory study that reports positive effect of a tangible VR system on spatial perspective taking ability.

ACKNOWLEDGEMENTS

The research has been supported by Canada's SSHRC Insight Grant program, the NSERC Discovery Grant program, and the

Canada Foundation for Innovation and the Ministry of Research and Innovation of Ontario. We also thank Dr. Dimitri Androutsos and the NSERC Undergraduate Student Research Award (USRA) program.

REFERENCES

[1] Alissa N. Antle and Sijie Wang. 2013. Comparing Motor-cognitive Strategies for Spatial Problem Solving with Tangible and Multi-touch Interfaces. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction* (TEI '13), 65–72. https://doi.org/10.1145/2460625.2460635

[2] Jack Shen-Kuen Chang, Georgina Yeboah, Alison Doucette, Paul Clifton, Michael Nitsche, Timothy Welsh, and Ali Mazalek. 2017. TASC: Combining Virtual Reality with Tangible and Embodied Interactions to Support Spatial Cognition. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (DIS '17), 1239–1251. https://doi.org/10.1145/3064663.3064675

[3] Paul G. Clifton, Jack Shen-Kuen Chang, Georgina Yeboah, Alison Doucette, Sanjay Chandrasekharan, Michael Nitsche, Timothy Welsh, and Ali Mazalek. 2016. Design of embodied interfaces for engaging spatial cognition. *Cognitive Research: Principles and Implications* 1, 1: 24. https://doi.org/10.1186/s41235-016-0032-5

[4] Rossana De Beni, Francesca Pazzaglia, and Simona Gardini. 2006. The role of mental rotation and age in spatial perspective-taking tasks: when age does not impair perspective-taking performance. *Applied Cognitive Psychology* 20, 6: 807–821. https://doi.org/10.1002/acp.1229

[5] Andreas Dünser, Karin Steinbügl, Hannes Kaufmann, and Judith Glück. 2006. Virtual and Augmented Reality As Spatial Ability Training Tools. In *Proceedings of the 7th ACM SIGCHI New Zealand Chapter's International Conference on Computer-human Interaction: Design Centered HCI* (CHINZ '06), 125–132. https://doi.org/10.1145/1152760.1152776

[6] John Eliot. 2002. About Spatial Intelligence: I. *Perceptual and Motor Skills* 94, 2: 479–486. https://doi.org/10.2466/pms.2002.94.2.479

[7] Augusto Esteves, Saskia Bakker, Alissa N. Antle, Aaron May, Jillian Warren, and Ian Oakley. 2015. The ATB Framework: Quantifying and Classifying Epistemic Strategies in Tangible Problem-Solving Tasks. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction* (TEI '15), 13–20. https://doi.org/10.1145/2677199.2680546

[8] Andrea Frick, Wenke Möhring, and Nora S. Newcombe. 2014. Picturing perspectives: development of perspective-taking abilities in 4- to 8-year-olds. *Frontiers in Psychology* 5. https://doi.org/10.3389/fpsyg.2014.00386

[9] Roland Guay. 1976. *Purdue Spatial Vizualization Test*. Educational testing service.

[10] M Hegarty, M Kozhevnikov, and D Waller. 2008. Perspective taking/spatial orientation test. *University California Santa Barbara. Retreived from: http://spatiallearning.org/resource-info/Spatial_Ability_Tests/PTSOT.pdf.*

[11] Mary Hegarty, Daniel R. Montello, Anthony E. Richardson, Toru Ishikawa, and Kristin Lovelace. 2006. Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence* 34, 2: 151–176. https://doi.org/10.1016/j.intell.2005.09.005

[12] Mary Hegarty and David Waller. 2004. A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence* 32, 2: 175–191. https://doi.org/10.1016/j.intell.2003.12.001

[13] Marja van den Heuvel-Panhuizen, Iliada Elia, and Alexander Robitzsch. 2015. Kindergartners' performance in two types of imaginary perspective-taking. *ZDM* 47, 3: 345–362. https://doi.org/10.1007/s11858-015-0677-4

[14] B Hommel, J Musseler, G Aschersleben, and W. Prinz. 2001. The Theory of Event Coding: a framework for perception and action planning. *Behavioral and Brain Sciences* 24: 849–878.

[15] Eva Hornecker and Jacob Buur. 2006. Getting a Grip on Tangible Interaction: A Framework on Physical Space and Social Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '06), 437–446. https://doi.org/10.1145/1124772.1124838

[16] Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. 2008. Reality-based Interaction: A Framework for post-WIMP Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08), 201–210. https://doi.org/10.1145/1357054.1357089

[17] Mi Jeong Kim and Mary Lou Maher. 2008. The impact of tangible user interfaces on spatial cognition during collaborative design. *Design Studies* 29, 3: 222–253. https://doi.org/10.1016/j.destud.2007.12.006

[18] Mi Jeong Kim and Mary Lou Maher. 2008. The Impact of Tangible User Interfaces on Designers' Spatial Cognition. *Human–Computer Interaction* 23, 2: 101–137. https://doi.org/10.1080/07370020802016415

[19] David Kirsh and Paul Maglio. 1994. On distinguishing epistemic from pragmatic action. *Cognitive Science* 18, 4: 513–549. https://doi.org/10.1016/0364-0213(94)90007-8

[20] Günther Knoblich and Rüdiger Flach. 2001. Predicting the effects of actions: Interactions of perception and action. *Psychological Science* 12, 6: 467–472.

[21] Günther Knoblich and Wolfgang Prinz. 2001. Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: Human perception and performance* 27, 2: 456.

[22] Keith R. Lohse, Courtney G. E. Hilderman, Katharine L. Cheung, Sandy Tatla, and H. F. Machiel Van der Loos. 2014. Virtual Reality Therapy for Adults Post-Stroke: A Systematic Review and Meta-Analysis Exploring Virtual Environments and Commercial Games in Therapy. *PLOS ONE* 9, 3: e93318. https://doi.org/10.1371/journal.pone.0093318

[23] C. Lorenzini, M. Carrozzino, C. Evangelista, F. Tecchia, M. Bergamasco, and A. Angeletaki. 2015. A Virtual Laboratory An immersive VR experience to spread ancient libraries heritage. In *2015 Digital Heritage*, 639–642. https://doi.org/10.1109/DigitalHeritage.2015.7419587

[24] Sandra C. Lozano, Bridgette Martin Hard, and Barbara Tversky. 2007. Putting action in perspective. *Cognition* 103, 3: 480–490. https://doi.org/10.1016/j.cognition.2006.04.010

[25] David Lubinski and Camilla Persson Benbow. 2006. Study of Mathematically Precocious Youth After 35 Years: Uncovering Antecedents for the Development of Math-Science Expertise. *Perspectives on Psychological Science* 1, 4: 316–345. https://doi.org/10.1111/j.1745-6916.2006.00019.x

[26] Ali Mazalek, Sanjay Chandrasekharan, Michael Nitsche, Tim Welsh, Paul Clifton, Andrew Quitmeyer, Firaz Peer, Friedrich Kirschner, and Dilip Athreya. 2011. I'M in the Game: Embodied Puppet Interface Improves Avatar Control. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction* (TEI '11), 129–136. https://doi.org/10.1145/1935701.1935727

[27] Maurice Merleau-Ponty. 1996. *Phenomenology of perception*. Motilal Banarsidass Publishe.

[28] Nora Newcombe. 1989. The Development of Spatial Perspective Taking. In *Advances in Child Development and Behavior*, Hayne W. Reese (ed.). JAI, 203–247. Retrieved November 27, 2016 from http://www.sciencedirect.com/science/article/pii/S0065240708604152

[29] Jean Piaget and Bärbel Inhelder. 1956. *The child's concept of space*. Routledge & Paul.

[30] Wolfgang Prinz. 2005. An Ideomotor Approach to Imitation. *Perspectives on imitation: Mechanisms of imitation and imitation in animals* 1: 141.

[31] J. Quarles, S. Lampotang, I. Fischler, P. Fishwick, and B. Lok. 2008. Tangible User Interfaces Compensate for Low Spatial Cognition. In *IEEE Symposium on 3D User Interfaces, 2008. 3DUI 2008*, 11–18. https://doi.org/10.1109/3DUI.2008.4476585

[32] Philip Robbins and Murat Aydede. 2009. *The Cambridge handbook of situated cognition*. Cambridge University Press Cambridge.

[33] Daniel L. Shea, David Lubinski, and Camilla P. Benbow. 2001. Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology* 93, 3: 604–614. https://doi.org/10.1037/0022-0663.93.3.604

[34] Barbara Tversky and Bridgette Martin Hard. 2009. Embodied and disembodied cognition: Spatial perspective-taking. *Cognition* 110, 1: 124–129. https://doi.org/10.1016/j.cognition.2008.10.008

[35] David H. Uttal, Nathaniel G. Meadow, Elizabeth Tipton, Linda L. Hand, Alison R. Alden, Christopher Warren, and Nora S. Newcombe. 2013. The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin* 139, 2: 352–402. https://doi.org/10.1037/a0028446

[36] Maria Virvou and George Katsionis. 2008. On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE. *Computers & Education* 50, 1: 154–178. https://doi.org/10.1016/j.compedu.2006.04.004

[37] Jonathan Wai, David Lubinski, and Camilla P. Benbow. 2009. Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology* 101, 4: 817–835. https://doi.org/10.1037/a0016127

[38] F. M. Williams-Bell, B. Kapralos, A. Hogue, B. M. Murphy, and E. J. Weckman. 2014. Using Serious Games and Virtual Simulation for Training in the Fire Service: A Review. *Fire Technology* 51, 3: 553–584. https://doi.org/10.1007/s10694-014-0398-1

[39] Margaret Wilson. 2002. Six views of embodied cognition. *Psychonomic bulletin & review* 9, 4: 625–636.

[40] Tom Ziemke. 2003. What's that thing called embodiment? In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Retrieved from http://www.ida.his.se/ tom/cogsci03.pdf