

Assessing navigation in real and virtual environments: a validation study

S T Koenig¹, G P Crucian², J C Dalrymple-Alford³, A Dünser⁴

^{1,4}HIT Lab NZ, University of Canterbury,
Private Bag 4800, Christchurch 8020, NEW ZEALAND

^{1,2,3}Department of Psychology, University of Canterbury,
Private Bag 4800, Christchurch 8020, NEW ZEALAND

¹sebas@virtualgamelab.com

^{1,4}www.hitlabnz.org, ^{1,2,3}www.psyc.canterbury.ac.nz

ABSTRACT

For navigation ability, a transfer of knowledge from virtual training environments to real-world scenarios has been shown in several studies in the past. The purpose of this investigation is to demonstrate the equivalence of a series of navigation tasks in complex real-world and virtual buildings. Instead of testing knowledge transfer in very simple environments, identical tasks are performed in either version of a complex building. 29 participants are shown twelve landmarks, followed by a battery of tasks which are carried out in the real building by half of the participants, whereas the other half performs identical tasks in a virtual model of the same environment. No significant differences or effects were found, but due to the multifaceted nature of the gathered data and large variability within groups, overlap of both groups' distributions was minimal. To discover the underlying factors of this variability, further research is needed. Usability and future development of virtual environments are discussed.

1. INTRODUCTION

Navigation is a highly complex skill of moving oneself, a craft or vehicle through novel and familiar environments. A variety of cognitive functions such as memory, visual and spatial perception and problem solving are involved in this important skill. Without the ability to navigate, humans are highly restricted in their independent living. If any of the involved cognitive functions is affected by a brain injury, the amelioration of navigation deficits is an important part of cognitive rehabilitation. However, navigation training during rehabilitation is often restricted to very few locations like the hospital or the patient's home. When faced with such limitations it is desirable to use simulations to retrain patients' lost abilities in a wide range of environments. Whenever people have to navigate through inaccessible or dangerous environments, the simulation of such environments for training purposes is of large benefit. Until now, knowledge transfer between real and virtual environments (VEs) and differences between several modes of knowledge acquisition have been studied. The results of these studies have been mixed.

In a study by Richardson, Montello and Hegarty (1999) participants learned the layout of a complex building either from a map, the real building or a VE similar to the real building. Test performance in the real building yielded significantly poorer performance by the VE group. After multiple training sessions within a large-scale virtual building Ruddle, Payne and Jones (1997) were able to demonstrate near-perfect route finding abilities of their participants which closely resembled the real-world performance of a similar experiment by Thorndyke and Hayes-Roth (1982). Koh, von Wiegand, Garnett, Durlach and Shinn-Cunningham (1999) compared real-world training with the participants' exposure to immersive, non-immersive visualizations and also an architectural 3D model of the same environment. During training participants were free to explore the environment to their liking. While the authors conclude that training in virtual and real space are comparable, no actual navigation behaviour was required during the testing phase and only estimations of bearings and distances were reported.

Taken together, many virtual navigation studies have been limited in several ways. They either involve learning a single predefined route or only encompass judgements of bearings and distances from stationary viewpoints. This type of learning is valuable when demonstrating training effects from VEs to the real world, but is not sufficient for quantifying how people navigate through their surroundings. It is even less appropriate for making predictions of real-world navigation behavior which is desirable in a clinical context.

Looking at predefined routes or knowledge of landmarks poses obvious restrictions on examining all of the behaviour which is involved in finding your way through a complex environment. Moreover, when people navigate in their daily life, their goals and priorities change often and unforeseen circumstances and obstacles arise, so that a single predefined route is not always a viable solution. Routes cannot always be rehearsed in advance and the navigator has to make inferences about alternative routes and the overall spatial layout of the environment. Assessing such configurational knowledge about the environment in addition to route knowledge is a step in the right direction. Witmer, Bailey and Knerr (1996) trained their participants in a complex office building and assessed route and configurational knowledge. However, their study is still limited to a predefined route and landmarks along that route. The authors' results suggest that using VEs for route learning is superior to maps, but inferior to real-world training. Configurational knowledge was unaffected by different learning media.

Examining navigation behaviour in all its complexity, this present study is explicitly comparing human navigation in a large real-world building and its virtual counterpart. The quantification of navigation behaviour is part of a planned framework for training, assessing and eventually predicting cognitive skills in a clinical context, with focus on patients with brain injuries. With such focus it becomes important to assess how and why people are getting lost, as it is the case with brain-injured individuals. Thus, an important aspect for this study's design is the high demand which is placed on the participants' navigation skills to provoke situations of temporary disorientation. The developed virtual reality simulation is intended for use in the day-to-day routine in rehabilitation settings. As such, usability, flexibility and compatibility with the needs of brain-injured individuals are of highest importance and are discussed in the context of the development cycle of such a VE. To satisfy the needs of clinicians, several navigation parameters which are relevant for predicting a patient's performance outside of the rehabilitation hospital are being analyzed. Lastly, it is hypothesized that those critical parameters of navigation ability - namely walked distance, received cues, amount of decision errors at intersections, distance estimations and pointing errors - are similar in complex real and virtual environments. The navigation time, the number of stops a participant makes and the total time spent standing still during navigation are predicted to be higher in the VE, as these variables are expected to be influenced by the VE interface.

2. METHODS

2.1 Participants

36 healthy, right-handed participants from the Christchurch community aged 40 or above and unfamiliar with the tested building volunteered for this study. Only 29 participants are included in the analyses as three participants withdrew from the study due to symptoms of simulator sickness, two participants were familiar with the tested environment and two participants were excluded due to missing data after a technical failure of a recording device. The specific age group was chosen to include users with a wide range of computer experience and to assess the age bracket of patients with stroke who are expected to be a primary target group of the developed navigation assessment in the future. Age of the participants ranged from 51 to 72 years in a real-building group while the age for a VE group ranged from 42 to 66 years. Male and female participants were equally assigned to both groups – six male and nine female participants in the real-building group and five male and nine female participants in the VE group. None of the participants reported playing computer games apart from non-spatial puzzles or card games.

2.2 Design

Participants were assigned to either a real-world or VE group in a randomized blocked design. Each participant was shown the same set of 12 target locations within the real version of a complex building on campus of the University of Canterbury, New Zealand. Following the initial learning phase, a series of pen and paper tasks for assessment of spatial abilities were completed. Finally, half of the participants returned to the real building to find the previously shown locations while the other half was asked to complete the same tasks in the virtual version of the campus building.

2.3 Materials

2.3.1 Real and virtual environment. The assessed environment is the seven-floor Erskine Building at the University of Canterbury, New Zealand. The building was chosen for its complexity and unusual layout with a combination of large open space and narrow corridors. Only the lower four of all seven floors were used for the assessment. A floor plan with targets for the first floor can be seen in the appendix. The four included floors differed substantially in regards to corridors, smaller staircases and landmarks. Several staircases throughout the building allowed for a large amount of possibilities to traverse from one landmark to the next.

The virtual model of the building was created using Google SketchUp 7 Pro. Textures were imported from photographs and floor plans were used to model the building to scale. Floor plans and measurements were displayed with Autodesk Design Review 2011. Most exit signs, direction signs and details within the building were included in the model with the goal of incorporating all constant features into the VE. Unfortunately, some minor aspects of the building change often so that not all details were up-to-date in the assessment. Interactions within the VE, data collection, interface and visual and navigation analysis tools were developed with the game engine Unity (version 2.6). The finished application records position and rotation of the participant within the VE in intervals of 100ms. The virtual environment was displayed using a three-screen back projection system with a field of view of 180 degrees. However, due to technical limitations only 120 degrees of the environment were displayed on screen so that the left and right edges on screen appeared slightly stretched. Each screen measured 2.44m x 1.83m. The participant was seated 2.2m in front of the center screen. This set up allowed the participants to show natural orientation behavior by turning to the side screens for searching the environment. The VE was rendered using a single PC with a quad-core CPU and three Nvidia GeForce GTX260 graphics cards running in SLI. All six projectors received input from two Matrox Triple-Head2Go graphics expansion modules. Only three of the six projectors were used to provide a non-stereoscopic rendering of the VE. This was done to allow the large interactive environment to run at a steady frame rate of 80 frames per second. Participants used a standard three-button computer mouse to navigate through the environment. Mouse movement was implemented to correspond to left-right and up-down head rotation and holding down right or left mouse buttons moved the participant back and forth at walking speed respectively. Sensitivity of the mouse was reduced substantially when compared to first-person computer games to allow participants with no computer experience to use the mouse accurately.

2.3.2 Pen and paper tests. Spatial abilities were measured with the Object Perspective Taking Test (OPTT; Hegarty & Waller, 2004), Mental Rotations Test (MRT; Vandenberg & Kuse, 1978) and the Card Rotations Test (CRT; Ekstrom, French, Harman & Dermen, 1976). In addition, orientation ability was assessed with the Santa Barbara Sense of Direction Scale (SBSODS; Hegarty, Richardson, Montello, Lovelace & Subbiah, 2002). Simulator sickness was assessed using the Simulator Sickness Questionnaire (SSQ; Kennedy, Drexler, Berbaum & Lilienthal, 1993). Computer experience was measured with an adapted version of the Computer/Internet Experience and Skills Questionnaire for: Internet Diabetes Trial at Harborview (Goldberg, 2004).

The OPTT is a test for spatial ability which requires judgements of bearings from imagined viewpoints. The test's score is the average of the absolute deviations from the correct angles over all of the test's twelve items. The MRT consists of three-dimensional line drawings (target) which need to be compared to four test objects. Two of the test objects are rotated versions of the target which need to be identified. The other two are mirror images or of slightly different structure. The test consists of 20 items, four test objects each, and the score is calculated by the number of attempted test objects divided by the number of correctly identified test objects. The CRT also requires the correct identification of test objects in comparison to a target object. However, the test and target objects are random polygons in different two-dimensional orientations. The number of attempted test objects divided by the number of correctly identified test objects is used as the test's score. Lastly, the SSQ assesses the severity of a range of symptoms of simulator sickness. This self-report measure states 16 symptoms for the domains of "nausea", "oculomotor" and "disorientation". The participant is asked to rate each symptom as either "none", "slight", "moderate" or "severe".

2.3.3 Navigation Test. Navigation through the Erskine Building consisted of two phases. During an initial learning phase, all participants were guided through the building on a predefined path which passed 12 target locations on four different floors. Four locations in the basement (Lecture Theater, Crypt Room, Bluebird Vending Machine, Dead End) and first floor (Main Entrance, Reboot Cafe, Coke Machine, Elevator) and two locations on second (Reception, Atrium Bridge) and third floor (Seminar Room, Walkway) were chosen. The total length of the walked training route is 498 meters. The walking distance was determined during a test trial in the VE which provided accurate measurements of the route. The lower four floors contain a total of 26 decision points where participants must choose between alternate paths. The basement contains three points, first floor contains four points, second floor contains eight points and third floor contains eleven points. Alternate paths are classified as either optimal, suboptimal or wrong. Unfortunately, there is no established way of classifying paths through a complex environment. With such a large amount of possible paths for each task, a simple dichotomy of correct or wrong paths does not capture all information. The following classification appeared to unambiguously classify each of the participant's decisions, no matter how complicated the taken path was.

- The optimal path is defined as the shortest single route which takes the participant from start to target.

- A wrong path is a decision which leads towards a wrong floor (i.e. target is up but participant goes down), along a route which does not lead to the target at all (i.e. dead end or wrong room) or any decision which is a direct turnaround on a path which leads optimally or sub-optimally to the target. That is, if the participant is standing very close to the target and has the optimal or suboptimal path ahead, but incorrectly turns around and retraces the previously walked path, this is considered wrong.
- All chaotic movement which cannot be classified as walking a defined path, but rather involves walking back and forth, in circles or any other diffuse behaviour, is considered an additional wrong decision.
- All remaining path choices are evaluated for the travel distance they require to reach the target, assuming that all following path choices minimize travel distance. The shortest of these paths is suboptimal, all others are wrong. If two or more are of the same shortest length, they are defined to be suboptimal. This is especially common in buildings with symmetrical layouts.

Optimal decisions were analyzed separately whereas suboptimal and wrong decisions were combined to an error score. Suboptimal decisions were scored as an error with a factor of one and wrong decisions were scored with a factor of two. The error score is the sum of all non-optimal decisions.

Half of the twelve target locations were secluded and allowed no direct line of sight to the other locations whereas the other half was in a more central location with higher visibility towards other locations and the layout of the building. This selection was based on the approach of Braaksma and Cook (1980) who evaluated the complexity of large buildings using a line-sight-network where visibility between all important landmarks within a building was assessed. This approach was applied to the seven-floor Erskine building to quantify the difficulty of each target location. However, the classification of target difficulty and its influence on navigation behavior is beyond the scope of this paper and will be discussed elsewhere. It was also not possible to control the order and amount of exposure that each location received during the initial learning route in such a complex environment. Though, the hidden locations naturally received less exposure whereas the central locations were seen more often during route traversal. Before walking along the learning route the participant was instructed to pay attention to the target locations and more importantly, to get a good sense of the overall layout of the building. Instructions also included the fact that the traversed learning route and order of target exposure were irrelevant for the following navigation test. Further, it was mentioned that all target locations were again to be rehearsed before starting the navigation test. Participants were not allowed to wander and had to stay with the experimenter at all times. Walking speed along the route was held constant as far as that was subjectively possible by the experimenter. Though, the participants were allowed to stop along the route and look around. Orientation behavior was strictly encouraged and initial instructions emphasized that the participant was free to do what he/she normally does when being in a novel environment. While questions were not forbidden, the only provided answer was the floor which the participant was currently on. Instructions and explanations about floors and target locations along the route were identical for each participant. Lastly, each target location was accompanied by a so-called reference point. In order to accurately point or navigate to the target during the test trial, an exact location was given to the participant (e.g. center of Coke Machine, middle of Elevator).

For the assessment phase of the study a different order of target locations was chosen to avoid the usage of the learning route. For the same reason the building was entered through a different entrance. Participants were expected to demonstrate configurational knowledge from the very beginning of the assessment by finding new ways through the building. Half of the target locations were designated navigation targets while the other half was used for pointing tasks. Order and nature of tasks was the same for all participants and both groups. Navigation and pointing tasks were always alternated.

For navigation tasks participants were instructed to find the shortest way to the given target without using elevators or asking other people for help. There were no route restrictions and all of the lower four floors were available for use by the participants. Cues were given systematically whenever participants asked for help or indicated that they were lost. Further, whenever a participant took more than 2 consequent wrong turns at a decision point or when no progress was made on a wrong floor (>4 decision points without leaving the floor towards the correct floor), a cue was given. Cues were categorized to either state that the participant is on the wrong floor, to verbally identify the correct floor, to give a semantic cue about the target (what the target looked like), to guide towards the correct side of the building, to name the reference point and to cue everything (explain in detail how to get to the target). Cues were given gradually in the listed order except when a participant asked specifically for a cue (e.g. "Which floor is the target on?"). Participants' navigation performance in the real building was recorded on video. All videos were later analyzed using VirtualDub to extract the timing of all tasks, cues, stops and to plot the exact route on a floor plan using Autodesk Design Review 2011. Performance in the VE was automatically recorded as described above. In addition, the plotted

paths on the floor plans were later rewalked in the VE with accurate timing and route to utilize the strengths of the VE to visualize and analyze the data. That is, by transferring all data into the VE, distances, number of stops, angles and viewpoints were easily computed, compared and displayed in 3D space with pinpoint accuracy. While measurements in the VE are accurate to about 2-3mm, transferring the video-recorded navigation data onto the floor plan and into the VE was done manually. Thus, location errors of up to 30cm and timing errors of up to half a second per stop are to be expected. For visual and computational analysis of the participants' routes, the original VE was modified using the Unity game engine to allow the experimenter to visualize all data, rewalk routes and carry out distance and angular calculations after the experiment was finished. The participant's route was plotted by a large number of data points which represented location and rotation data within the VE at a given time. Data points were rendered as simple sphere objects in the VE and were identified by a time stamp to follow the participant's route visually and chronologically.

Pointing tasks were given whenever the participant reached a navigation target. As soon as a target was reached, a tripod with an attached protractor was set up at a predefined location. The tripod had a wooden plate mounted on top with a clock-hand attached to it. The protractor was hidden underneath the wooden plate to prevent giving any cues to the participant. The clock-hand was used by the participant to indicate the direction in which the pointing target was expected to be. The protractor indicated the angle in which the participant pointed. The absolute deviation from the correct angle was recorded. Pointing targets were always on the same floor and not visible from the participant's position. Participants were not allowed to leave the location where the tripod was set up. After pointing towards the target, the participant was asked to estimate the egocentric euclidian distance (in a straight line) towards the pointing target. The participant's answer was scored as a percentage of the actual target distance. Next, the difficulty of the last navigation and pointing task were assessed. The participant indicated the difficulty of both tasks on two continuous Likert scales with anchor points of "too easy" and "too difficult". Difficulty was measured as the percentage of the participant's mark on the Likert scale (i.e. 0-100). Lastly, an empty floor plan (only the outer walls of the current floor of the building) was provided in which the participant had to draw his current position where the tripod was set up and also the position of the pointing target. The floor plan was aligned with the participant's view and the location of two landmarks (main entrance and side entrance) was shown on the floor plan to give the participant a better sense of distance and location. For the purpose of analyzing the participant's answers, the floorplan was divided into a three by three array of sections as seen in the appendix. The different sized sections were chosen to match the irregular layout of the building. The deviation from the correct section was counted so that diagonal movement was not allowed. The highest possible error score for any response is therefore four, given that the participant's mark is in the opposite corner of the building from where the correct target location is.

2.4 Procedure

Participants were tested individually with each session lasting between two and three hours. The experiment started at the Psychology Building at the University of Canterbury, New Zealand. After an initial briefing and additional information about simulator sickness, participants gave written informed consent. During the first 15 minutes questionnaires for demographic background, computer experience and the SBSODS were completed. Next, the participant was taken to the Erskine Building. The experimenter led the participant along the predefined training route through the building and explained all twelve targets. After leaving the training environment, the participants returned to the Psychology Building where they completed the MRT, OPTT and CRT within about 30 minutes. Before navigation performance was tested, a list of the twelve targets in random order was presented. The participant had to recall as many details about the target locations as possible. Such information included context information (what happened near the object), floor number, side of the building, location of the reference point or how to get to the target. All answers were recorded, scored for their correctness (each item either as correctly or incorrectly remembered), and the participant received feedback and further explanations about all targets. If the participant felt confident about all target locations and no questions remained, the navigation assessment started. Half of the participants were guided towards a different side entrance of the Erskine building where the assessment began. The other half of the participants were tested at the HIT Lab New Zealand, where the VisionSpace back projection theater is set up. Initially, participants were asked to report any symptoms of simulator sickness using the SSQ and received an introduction to using the mouse for navigation through the VE. A simple open environment with two visible targets was used as a practice scenario for navigation and pointing tasks. After each participant was comfortable with using the mouse, the virtual assessment started at the exact same virtual side entrance which was also used for the real-world assessment. The remainder of the testing session was identical for both groups as has been described previously. The SSQ which participants had to fill out again after their exposure to the VE was the only additional questionnaire for this group. Before and after scores were compared and analyzed.

2.5 Statistical Analysis

The basis of the present study is a comparison of real and virtual navigation with a prediction of equivalence being made between both groups. This by itself poses a problem, because most statistical procedures are designed to test for differences between groups. Further, the absence of a significant difference is not to be interpreted as both groups being equivalent (Nickerson, 2000). Any difference between two or more groups can potentially be shown to be statistically significant, given a large enough sample size. Thus, a different method for interpreting the equivalence of groups is needed. Tryon (2001) and Tryon and Lewis (2008) suggest the use of a range (delta) which is defined by the extreme points of adjusted inferential confidence intervals (ICIs) of both groups. If delta is smaller than a predefined range of indifference, statistical equivalence is established. However, the range of indifference needs to be determined on substantive grounds which is not a trivial task for any research question.

In conclusion, a hybrid approach to statistical analysis has been chosen. Firstly, both groups are compared with standard statistical tests (i.e. t-tests for independent means) seeking to find a significant difference. Effect sizes for all comparisons are also calculated following the procedures of Cohen (1987). Lastly, the ICIs for both groups are calculated and their overlap is determined. The overlap measure indicates the percentage of participants which are included in the overlapping area of both groups' ICIs.

3. RESULTS

Levene's tests for homogeneity of variances have been conducted for all comparisons of navigation performance and pen and paper tests. The variances for Total Time of Stops differ significantly between the VE and real building group ($F(1,27) = 14.889, p = 0.0006$). No other Levene's test shows a significant difference. Lilliefors' tests for normality indicate that none of the distributions reported in this study differ significantly from a normal distribution.

As predicted, most critical navigation parameters did not show a significant difference at $\alpha = 0.05$ (see table 1). Performance in the pen and paper tests was not significantly different between the real-world and VE group. There was no effect evident for the CRT ($d=0.19$). MRT and OPTT showed small effect sizes of $d=0.48$ and $d=0.41$ respectively.

For navigation distance no significant difference was observed. Participants in the VE group on average travelled 46 meters further than participants in the real environment. Cohen's d was found to be small for this comparison ($d=0.37$). Confirming our expectations, there was a significant difference between navigation time of both groups. Participants spent significantly more time navigating through the virtual Erskine Building which results in a large effect size ($d=1.76$). Variance for the VE group was especially large and individual performance ranged from 501s to 2111s. Navigation time was based on the total time for all six navigation tasks from the moment the participants started walking along the first path until they found the last navigation target. Pointing tasks were not included in this measure.

The number of cues that were given to the participants did not differ significantly between both groups. However, a medium effect size ($d=0.6$) indicates that there was a difference in the number of help cues the VE group and the real-building group received even though this difference did not reach significance. This might also be due to the large variance in the VE group where two participants received 20 and 21 cues respectively. When both participants are excluded from the analysis, the effect size is reduced to 0.24. A t-test for independent means shows no significant differences between both groups when the outliers are removed ($t(25)=-0.622, p=0.539$).

The comparison of navigation decisions is of central importance, as this measure directly quantifies how participants navigated through the real and virtual environment. As predicted, no significant differences were found for the number of optimal decisions and the decision error score. No effect was found for the number of optimal decisions ($d=0.17$) and a small effect was evident for the decision error score ($d=0.28$). Systematic errors when drawing navigation and pointing targets onto empty floor plans showed no significant difference. The effect size for this comparison was found to be small ($d=0.22$). No significant difference was found for distance estimations in both groups. Nonetheless, participants who were assessed in the VE appeared to consistently underestimate the true distances towards pointing targets ($d=0.66$). The largest estimation errors occurred when judging the shortest distance. For the distance between Elevator and Main Entrance (6.97 meters) average judgements were 156.9% and 172.96% of the true distance in real building and VE respectively. This was the only overestimated pointing target for the VE group across all tasks. A large effect size for angular pointing errors ($d=0.73$) indicates that pointing errors were larger in the VE than in the real building. The difference between both groups is non-significant. The remaining comparisons for number of

stops and total time spent standing still both show large effect sizes ($d=1.61$ and $d=1.67$) and significant differences between both groups. Participants navigating through the VE stopped significantly more often and spent more time without virtual movement.

Table 1. Means, Standard Deviations, P-Values, Effect Sizes and Confidence Interval overlap for all navigation measures.

Measure	Mean±S.D. Real Group	Mean±S.D. VE Group	df	P-value	Effect Size d	CI overlap %
CRT	0.892±0.087	0.906±0.057	27	0.6037	0.19	23
MRT	0.822±0.065	0.851±0.059	26	0.2231	0.48	13
OPTT	38.156±23.832	30.224±14.796	27	0.2953	0.41	6
Navigation Distance	443.612±132.922	490.096±118.087	27	0.3296	0.37	24
Navigation Time	674.733±248.775	1380.5±520.627	27	0.0001*	1.76	0
Number of Cues	4.867±3.681	7.857±6.323	27	0.1281	0.6	3
Decisions Error Score	23.14±12.28	26.64±12.88	27	0.4684	0.28	21
Optimal Decisions	12.86±2.96	13.36±2.76	27	0.6477	0.17	29
Floor Plan Errors	6.73±3.37	7.57±4.29	27	0.5620	0.22	38
Distance Estimation	120.01±50.91	90.42±38.4	26	0.0944	0.66	3
Angular Pointing Errors	30.983±19.354	45.228±19.848	27	0.0608	0.73	0
Number of Stops	17.733±10.43	40.357±17.627	27	0.0002*	1.61	0
Total Time of Stops	171.4±111.617	589.746±388.33	27	0.0004*	1.67	0

Note: * indicates a significant difference at $p < 0.05$; CRT – Card Rotations Test; MRT – Mental Rotations Test; OPPT – Object Perspective Taking Test

In addition to the aforementioned analyses, inferential confidence intervals (ICIs) were calculated for both groups of all measures (Tryon, 2001; Tryon & Lewis, 2008). This was done to further investigate the equivalence of the two groups. By simply looking at the results of a standard t-test it is impossible to conclude that two groups are the same. The amount of overlap, that is the number of data points in the overlapping range of both ICIs, is an indication for the equivalence of both groups. However, as a result of this analysis almost no overlap was evident (see table 1). Floor plan errors showed the highest overlap with eleven of the 29 participants in the overlapping range of the two groups' ICIs (38%). All remaining participants were located to the extreme left and right of the distribution of error scores. As expected no overlap was found for variables which also showed a significant difference.

Correlations of pen and paper tests (MRT, CRT, OPPT) with our navigation parameters were non-significant throughout. Only CRT score and errors in the floor plan task correlated significantly ($r=-0.516$, $p=0.007$) such that higher CRT scores are associated with less errors in this task. Also, age and sex showed no significant relationship towards any of the navigation measures.

Computer experience of the participants in the VE group was correlated with all navigation outcome measures. Correlations were generally negative and non-significant. Computer experience and the number of optimal decisions along the traversed route are positively correlated ($r=0.564$, $p<0.05$) which suggests that participants with higher computer experience performed better in the VE.

The participant's experience with the VE was almost entirely positive. Those participants who were unaffected by simulator sickness enjoyed the test session very much. While few participants suffered from mild symptoms of simulator sickness and three participants had to withdraw from the study due to more severe symptoms, the average increase of the total score from pre-assessment to post-assessment was 32.21 (SD=40.37) over 18 participants. The three participants who withdrew from the study reported total scores of 134.6 (two of them) and 120. Five of the participants showed no change in their score or even lower symptoms compared to the initial SSQ assessment.

4. DISCUSSION

In this pilot trial, validity of performance measures in a complex virtual building was assessed and navigation performance in the VE and its real-world counterpart were directly compared. The navigation tasks focused on configurational knowledge of the building and the 29 participants were required to make inferences about the shortest routes which had not been part of their previously shown learning route.

Most navigation parameters did not show a significant difference between the real-building and VE group. When participants were required to make decisions along their travelled routes, their decision errors

and their choices for the optimal, shortest route did not differ significantly between groups. Both variables are considered the most comprehensive indicators of navigation performance, because they capture most of the participants' behaviour. In addition to the standard statistical analyses, effect sizes were calculated in order to further support the hypotheses of equivalence of both groups. The small to very small effects for both variables supported the initial results. Another variable of importance is the number of cues which the participants received to find the navigation targets. Again, both groups did not differ significantly and a medium effect size was observed. Only after removing data of two participants who received most of the cues in the VE group, the effect was reduced to small size. Both participants had difficulties adjusting to the VE and using the navigational interface. Due to their difficulties to navigate adequately, abrupt viewpoint changes resulted in symptoms of simulator sickness so that breaks between navigation tasks were needed. Consequently, the removal of data points from this analysis seems justified.

A variety of other measures were used to quantify the participants' ability to find their way through the building and estimate the position of targets around them. None of these measures produced a significant difference, but effect sizes varied considerably between tasks. Distances in the VE were consistently underestimated which is in line with many previous findings in the literature (e.g. Witmer & Kline, 1998; Koh et al., 1999; Thompson et al., 2004). However, contrary to most experiments, targets were not visible from the participant's viewpoint and had to be judged based on configurational knowledge of the building rather than visual cues. An exception to the underestimations was apparent for the shortest true distance where participants in the VE on average overestimated distance by 73%. The participants' viewpoint for this task was mostly occluded by a virtual wall which differs to all other distance judgements where no occlusions were present. Thus, participants had to rely solely on their mental map of the building layout instead of using visual cues of the VE for their judgement. The resulting difference between both groups for this particular pointing task was very small and no effect was evident. This suggests that participants' beliefs about distances are naturally accurate or too long, but are contradicted by exposure to visual cues in the VE.

Angular pointing errors towards unseen targets differed substantially between groups which is shown by a medium effect. This difference is expected to be of technical nature. In the VE, participants carried out the pointing task by positioning a red dot in the desired direction. This red dot was rendered in the center of the three-screen display and was explained to be similar to an index finger for pointing in a direction. A tripod with a clock-hand was used in the real building which provided a top-down, exocentric way of pointing towards a direction. While the former method for pointing intuitively seems more accurate, the VE group made larger absolute errors for their bearing estimations. One possible reason for these errors is the distortion of the displayed VE. Displayed field of view and screen resolution did not match so that a distortion of the VE was visible on the side screens. This technical limitation needs to be addressed in future versions of navigation assessments to provide a better user experience and results of higher validity.

The number of stops and the total time participants stopped on their routes were intended to assess the extent to which each person showed orientation behavior. Participants in the real building used such stops to look around, search for landmarks and find their bearings. Unfortunately, many additional stops in the VE were due to difficulties with the user interface. Participants slowly adjusted to the navigation with a computer mouse and several additional stops were recorded when participants needed to shift the mouse to the middle of the table, because they were running out of space. Further, several users adopted a strategy of repeatedly pushing the left mouse button instead of holding it down for forward movement as intended. Many additional pseudo-stops were also recorded when participants walked up and down stairs. The collision detection with stairs was not as smooth as with other parts of the virtual model and needs refinement. Consequently, the computer-based analysis of stopping behavior was programmed to eliminate such "micro-stops" by setting a threshold of one second for the shortest possible stop. With these limitations in mind, it comes as no surprise that a significant difference for both variables of stopping behavior was observed and the results of these analyses cannot contribute to the interpretation of navigation ability as intended.

To further substantiate our hypotheses of equivalence for all navigation variables, an additional analysis was conducted which uses the amount of overlap of inferential confidence intervals (ICIs). General overlap between scores of both groups was very low for all variables. The analysis revealed that a substantial amount of data points were located at the extreme positive and negative ends of the parameters' distributions. The finding of such small overlap of our groups in light of no significant differences and small effect sizes suggests that further research needs to be conducted to explain navigation behavior in very complex natural environments.

During the course of this study it became apparent that several usability issues had to be addressed for the future use of VEs. In a series of preliminary usability tests, navigation interfaces were evaluated for their ease of use. Main criteria for using a three-button computer mouse were the widespread use of the device and little

physical effort needed for handling the device. Both are especially important for single-handed use when the other hand is unavailable, for example in cases of hemiplegia. To address the issue of simulator sickness, further refinements need to be made to the projection of the VE. For future studies the separate rendering of three viewports, one on each of the three screens of the VisionSpace theater, solves the problem of perspective distortion. This solution was not viable for the present study as the virtual building was too large for three separate renderings. For clinical use of the application head-mounted displays and portable projectors are considered in future trials.

During testing sessions the accurate modeling of smaller details emerged as an important aspect of the VE. Several participants reported using posters on the corridor walls for orientation. Unfortunately, due to time-restraints most posters and several other non-permanent minor details had not been included in the virtual model of the Erskine building. These limitations can easily be addressed with a more efficient work flow which optimizes the integration of Google SketchUp and Unity in the development cycle of a VE. The integration of this workflow into clinical day-to-day routine at a rehabilitation hospital will be tested in upcoming trials.

In order to consider the use of VEs in cognitive rehabilitation the difficulty of three-dimensional environments has to be evaluated. Such quantification of difficulty is necessary to provide alternate versions of navigation tasks, to classify routes and environments which patients are exposed to, and to adjust training difficulty in the context of a comprehensive rehabilitation training. In recent experiments, researchers manipulated the number of turns or the length of the route, because they are simple to measure and implement in an experimental setting. However, most real-world environments cannot be compared to the average office corridor in a university building. Residential houses, shopping malls, train stations or many other places often have multiple floors and there is more than one viable path which leads to the target. For simulation of these scenarios different measures need to be found in order to assess and score complex behavior in a standardized, systematic way. Braaksma and Cook (1980) propose a line-of-sight approach for landmarks in complex buildings like transportation terminals. How visibility, number of possible routes, results of pathfinding algorithms or other yet undefined variables influence the navigation performance in complex environments must be subject to further investigation. In addition, the relationship of VE performance and well-established clinical measures of spatial abilities needs to be examined. The results of Nadolne and Stringer (2001) and Kozhevnikov and Hegarty (2001) suggest that small-scale tasks like mental rotation place different demands on the cognitive system than navigating through the environment. Hence, more ecologically valid assessments are needed and the continued evaluation of VEs for such purpose seems justified.

In conclusion, the VE assessment has proven to be a useful tool for accurately capturing a complex skill like navigation ability. Through the strengths of virtual reality environments the capture, interpretation and visualization of navigation data has been greatly improved. However, our results show no correlations with other measures of spatial ability and the complexity and high variability of our data did not allow for an unambiguous interpretation. This suggests that measuring navigation ability in all its facets is a highly complex matter which cannot easily be related to existing measures of configurational knowledge of environments. To further increase the validity of gathered navigation data, several improvements towards higher usability of the VE are necessary. Issues of simulator sickness, display distortion and model detail of the VE need to be addressed. With refined navigation measures and larger sample sizes more insights into the underlying factors of navigation performance variability are expected. Such insights are especially needed to utilize more ecologically valid assessments, because with higher ecological validity complexity of the assessment increases substantially.

5. REFERENCES

- J Cohen (1987), *Statistical power analysis for the behavioral sciences* (rev. ed), Lawrence Erlbaum Associates, Hillsdale, N.J.
- R B Ekstrom, J W French, H H Harman and D Dermen, *Kit of factor-referenced cognitive tests*, Educational Testing Service, Princeton, NJ, 1976.
- H Goldberg (2004), *Computer/Internet Experience Skills Questionnaire for: Internet Diabetes Trial at Harborview*, Retrieved June 26, 2010, from http://www.hetinitiative.org/grantee_instruments/UwashHart_Baseline%20Computer%20Skills%20Survey.pdf.
- M Hegarty, A E Richardson, D R Montello, K Lovelace and I Subbiah (2002), Development of a self-report measure of environmental spatial ability, *Intelligence*, **30**, pp. 425-447.

- M Hegarty and D Waller (2004), A dissociation between mental rotation and perspective-taking spatial abilities, *Intelligence*, **32**, pp. 175-191.
- R S Kennedy, J M Drexler, K S Berbaum and M G Lilienthal (1993), Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness, *International Journal of Aviation Psychology*, **3**, 3, pp. 203-220.
- M Kozhevnikov and M Hegarty (2001), A dissociation between object-manipulation spatial ability and spatial orientation ability, *Memory and Cognition*, **29**, pp. 745-756.
- M J Nadolne and A Y Stringer (2001), Ecologic validity in neuropsychological assessment: Prediction of wayfinding, *Journal of the International Neuropsychological Society*, **7**, pp. 675-682.
- R S Nickerson (2000), Null hypothesis significance testing: A review of an old and continuing controversy, *Psychological Methods*, **5**, 2, pp. 241-301.
- A E Richardson, D R Montello and M Hegarty (1997), Spatial knowledge acquisition from maps and from navigation in real and virtual environments, *Memory and Cognition*, **27**, 4, pp. 741-750.
- R A Ruddle, S J Payne and D M Jones (1997), Navigating buildings in “Desk-Top” virtual environments: Experimental investigations using extended navigational experience, *Journal of Experimental Psychology: Applied*, **3**, 2, pp. 143-159.
- P W Thorndyke and B Hayes-Roth (1982), Differences in spatial knowledge acquired from maps and navigation, *Cognitive Psychology*, **14**, pp. 560-589.
- W W Tryon (2001), Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests, *Psychological Methods*, **6**, 4, pp. 371-386.
- W W Tryon and C Lewis (2008), An inferential confidence interval method of establishing equivalence that corrects Tryon’s (2001) reduction factor, *Psychological Methods*, **13**, 3, pp. 272-277.
- S G Vandenberg and A R Kuse (1978), Mental rotations, a group test of three-dimensional spatial visualization, *Perceptual and Motor Skills*, **47**, pp. 599-604.
- B G Witmer, J H Bailey and B W Knerr (1996), Virtual spaces and real world places: transfer of route knowledge, *Int. J Human-Computer Studies*, **45**, pp. 413-428.

