Evaluating the Impact of Spatial Ability in Virtual and Real World Environments

Georgi Batinov, Kofi Whitney, Les Miller, Sarah Nusser and Bryan Stanfill
Iowa State University
Ames, Iowa, USA
{batinov,kwhitney,lmiller,nusser,stanfill}@iastate.edu

Kathleen T. Ashenfelter
US Census Bureau
Washington, D.C. 20233
kathleen.t.ashenfelter@census.gov

***Abstract*— Survey agencies in the United States continue to move many map-based surveys from paper to handheld computers. With large highly diverse workforces, it is necessary to test software with a diverse population. The present work examines the performance of participants grouped by their level of spatial visualization. The participants were tested in either the field or in a fully immersive virtual environment. The methodology of the study is explained. The performance of the participants in the two environments is modeled with least squares regression. Results of the study are presented and discussed.**

Keywords- map-based survey; virtual reality; spatial ability

I. INTRODUCTION

Survey agencies in the United States have been moving towards using handheld computers to replace the use of paper in their field operations. Since most field surveys are inherently location dependent, a lot of the software used will be map-based. Agencies, like the Bureau of Census, are forced to couple this move to map-based software with a highly diverse workforce, especially in large scale operations like the decennial census. Due to the wide range of individual differences typically encountered in such diverse workforces, software testing is a critical component of this process. Ultimately, the software has to be tested in the field to fully understand how it will perform. However, a significant issue with testing in the field is the cost. An interesting question is the viability of doing at least the initial testing of software in a virtual environment.

In the present work a study is described that looks at the participant's performance in either the field or in a fully immersive virtual environment. The task chosen for the study was address verification, where a census worker is given a list of addresses and they are expected to either determine the address is correctly located on their map or make the necessary corrections. The contribution of this paper is the direct comparison of a complex real world operation performed in both the field and virtual reality. This initial study didn't show very much difference between performance in the field and virtual environments. We did see a significant impact of the role that spatial visualization played in both environments.

In the remainder of the paper we look at related work, examine the methodology used in the study, present the statistical results, and provide a discussion of what we found.

II. BACKGROUND

Wobbrock et al. [24] proposed ability-based design as a paradigm for constructing individual-centric systems. According to Murray & Kluckhohn [18],"Every man is in certain respects (a) like all other men, (b) like some other men, (c) like no other man" (p.35). Benyon, Crerar, & Wilkinson [3] predicated the prominence of cognitive differences in human-computer interaction on the divide between physical and digital artifacts and noted that cognitive differences may have amplified effects in computing contexts (pp. 21-22).

Spatial ability is a compound factor that has often been linked to performance in interactive tasks. Several authors have used factor-analytic techniques to decompose spatial ability into constituents. In the nineties, [5] and [16] reported that it consists of five parts: spatial visualization, speeded rotation, closure speed, closure flexibility, and perceptual speed. Earlier publications by [6] and [20] suggested other combinations. Spatial visualization ability–defined by [9] (p. 173) as "the ability to manipulate or transform the image of spatial patterns into other arrangements"–has been shown to correlate with performance with command-line interfaces [11, 4], file system navigation [22], searching an information retrieval system [8], web browsing [25], simulated driving [1], and remote control of robots [15].

Beyond the combination of software and hardware, we also need to consider the field operating environment, which presents a multitude of stimuli and a continuously changing external context, unlike the traditional computer desktop. Whether a field setting can be reasonably approximated in a laboratory virtual environment is still an open question. Ref. [17] highlights the tradeoff between experimental control and ecological validity in traditional research methods and suggest that improved-fidelity virtual reality may reduce the compromise. If correct, their claim has practical implications, as well: virtual reality may become a low-cost alternative for field training. Two components that distinguish reality from a virtual environment are distance perception [23] and embodiment [2]. Ref. [10] compared environmental learning from the real world, non-interactive

 ISBN: 978-1-61208-250-9

video, and a desktop virtual environment, concluding that spatial ability is correlated with learning in both the real world and virtual environment, with a stronger effect for the desktop simulation. Ref. [21] compared walking patterns in reality and on a treadmill in non-immersive virtual setting, noting persisting small differences in gait after 20 minutes of acclimatization. Ref. [19] showed that increasing display size and resolution improved wayfinding and object location performance in a non-immersive virtual setting. The outcomes in these publications suggest that as we improve visualization and locomotion technologies, we may be able to run virtual reality experiments approaching ecological validity. Refs. [13] and [14] used a study design that appears similar to ours but was driven by a different agenda. The authors constructed a high-fidelity virtual reality model of a residential area in the United Kingdom and asked 27 participants to navigate to five locations inside an immersive environment designed by [7]. Participants had access to schematic maps, detailed maps, and written and spoken route instructions on a handheld device that served as a pathfinding aid. The authors described three different pathfinding behaviors, noted geographical hot spots for handheld device activity and per-destination aggregate device activity. The focus of the research was on linking location to handheld usage, and the authors did not report measures of statistical validity. In contrast, we set out to find statistical evidence for performance differentials on a software map task. Our experiment includes both a virtual reality and a field setting. Additionally, our task is a software task that has a navigation component. Finally, the virtual environment in [13, 14] had a lesser degree of immersion, because participants navigated with a joystick and the virtual model was projected on up to three VR walls.

III. MATERIALS AND METHOD

The experiment contained a screening phase and an exercise phase.

A. Screening phase

During the screening phase, one-hundred-and-twenty-four participants were individually assessed on spatial visualization, visual memory, perceptual speed, and perspective-taking ability. The tests were VZ-2, MV-2, and P-2 given in [9], and the perspective-taking assessment described in [12]. Participants with spatial visualization scores greater than or equal to 15 or less than 9 (out of 20) were randomly assigned to one of two treatments in the exercise phase. Pairs from either the low or high spatial visualization groups were randomized together, allowing each participant a 0.5 probability of assignment to either be tested in the virtual reality treatment or the field treatment. Thirty-two participants (14 males, 18 females) were assigned to the second phase of the experiment.

B. Field phase

For the field treatment, 15 participants (8 males, 7 females) were taken individually to the same spot in a residential neighborhood in Ames, Iowa. They were first trained on using the handheld device, locating addresses in the field, and the think-aloud protocol. An observer provided them with a stylus and a handheld computer: a Pharos Traveler 535x with a 240x320, 3.5” transflective screen and a 624 MHz Intel PXA270 processor. The observer explained the address verification task.

Participants would have to physically walk to an address in order to answer the question. If the map contained errors, they had to use the software’s editing features to position the address at the correct location or remove it altogether. Four outcomes were possible. An address needed to either be added to the map, deleted, moved to a new location, or confirmed without changing the map. Participants were told to only correct the addresses in their task list and to ignore other possible errors on the map. (The map contained no errors outside of scenario addresses.) Participants were then taught how to edit the software map and were also instructed to verbalize all their thoughts for a think-aloud protocol. The map software was started in training mode and participants were asked to locate and verify three training addresses in the immediate vicinity, while the observers answered procedure questions and provided feedback on the quality of the think-aloud. At the end of the training session, observers answered the participant’s final questions, and also explained that observers would not talk during the actual exercise, other than to prompt the participant to keep verbalizing or to ask about behavioral details. Observers then returned the participant to the location where all trainees started, switched the map software to experiment mode, and started an audio recorder (worn by the participant) and a GPS tracker (carried by the observer). The GPS tracker was not given to the participant so that they would not be interrupted to time-stamp address completions.

All participants verified the same six addresses off of an identical randomized list order (Figure 2), and therefore could not benefit from completion sequence hints. The list could be viewed at all times in the software by tapping the currently selected address. Errors for each address in the task list are shown in Figure 1. Participants were allowed to work on addresses in any order and could return to previously submitted addresses as many times as they wanted. Only final answers were evaluated for correctness. When finished, participants were then taken to two locations on the map and asked to point in the direction of the starting spot. Finally, observers audio-recorded an exit questionnaire detailing the participant’s perceptions of the study.

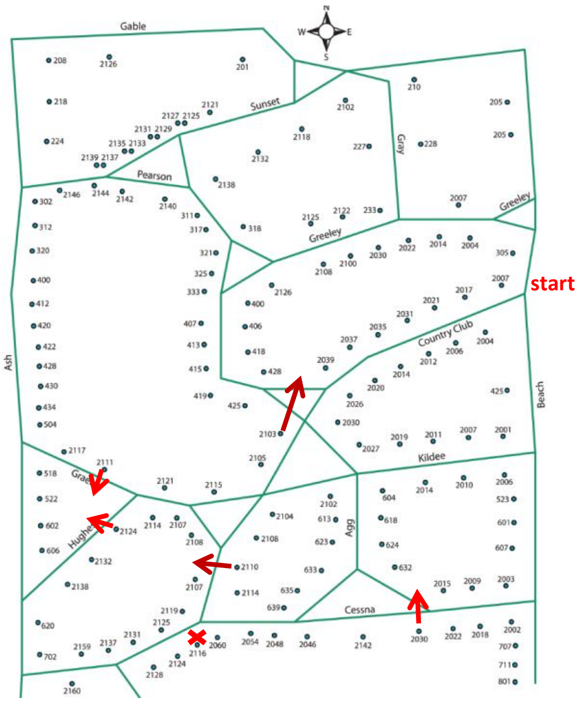

Figure 1. Address errors introduced on the map.

C. Virtual reality phase

Seventeen participants (6 males, 11 females) were randomly assigned to the virtual reality treatment and were taken individually to the VRAC C6 immersive virtual reality environment on the Iowa State University campus.

Virtual reality model-The virtual setting loaded in the environment was a high-fidelity three-dimensional model of the residential area portrayed in Figure 1, with an additional block modeled outside the westernmost and easternmost extents of the map. The model was created in SketchUp (www.sketchup.com) and imported into the virtual reality environment through VR Juggler (www.vrjuggler.org). Housing units and streets were georeferenced. However, actual housing units were represented by house models of similar size and style selected from Sketchup's repository of three-dimensional housing models (sketchup.google.com/3dwarehouse/). The neighborhood model also incorporated notable landmarks in the area, such as, street signs, curbs, textured surfaces, a day sky with sun, trees, shrubs, a playground, and a large building on the Iowa State University campus that was visible in some parts of the study area. The model did not include sidewalks, but did represent multi-lane streets and split boulevards, keeping throughway widths consistent with reality.

Virtual reality equipment-The virtual reality room is a cube with dimensions 3.05 x 3.05 x 3.05 m. Each of the four walls, floor, and ceiling displayed stereo images of 4096 x 4096 pixels at approximately 16 frames per second. Video projection is driven by a cluster of 48 HP xw9300 workstations with 96 nVidia Quadro graphics cards sending video frames to 24 Sony SRX-S105 digital cinema projectors. InterSense's IS-900 tracking system tracked the participant's head location and gaze direction, and the stereo perspective dynamically shifted with the user's gaze. The participants wore active stereo glasses.

Moving in virtual reality-Movement in the environment was accomplished by stepping towards the desired direction. A circular spot in the center of the floor, approximately 0.6 m in diameter, was the "dead zone". If the participant's head was located in the column of the spot, all movement stopped. Stepping outside the dead zone would start moving the virtual reality model in the opposite direction of the step, giving the illusion of the participant moving through the model in the direction of the step. As the participant stepped closer to the walls, movement speed increased, from approximately 0.1 m/s to a maximum of approximately 2.22 m/s (8 km/h or 5 mi/h). We fixed the maximum speed to a slow trot, because we were concerned that a higher speed could not be encountered in the range of walking speeds available to participants in the field treatment, and a lower maximum speed might bore participants, causing them to lose focus.

Protocol changes-Study protocol was exactly the same as in the field treatment, but prior to introducing the handheld device, participants were trained on moving inside the virtual environment. Participants also started training and the exercise at the same geographic spot in the virtual model as participants in the field.

Data collection and analysis-We tracked: distance traveled via GPS and virtual movement logs, time taken to complete the task, and number of addresses incorrectly verified (number of task errors). The virtual reality model was georeferenced, so travel coordinates within the immersive environment reflected actual distances. Additionally, we recorded all handheld software actions and user speech from the end of the training session to the end of the exit questionnaire.

We used least squares regression to explore statistical relationships among the data. Our response variables were distance traveled, time taken to complete the task, and number of errors. Predictor variables included spatial visualization category (low or high); field/virtual environment category; gender; perceptual speed, visual memory, and perspective-taking scores; and zoom, pan and map reset actions.

We hypothesized that:

> **Hypothesis 1:** High-spatial-visualization participants would travel significantly shorter distances than low-spatial-

visualization participants in both the field and virtual environments.

Hypothesis 2: High-spatial-visualization participants would take significantly less time than low-spatial-visualization participants in both the field and virtual environments.

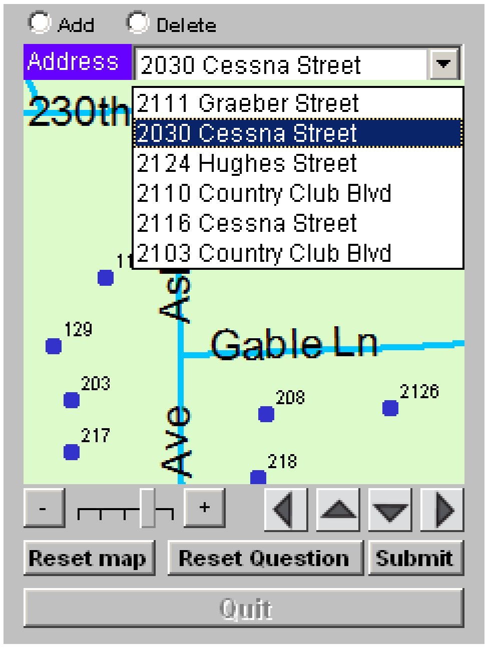

Figure 2. Edit screen with address list extended.

Figure 3. Participant in the virtual reality treatment.

In addition, we expected that there would be some impact on the participants between treatments, especially the participants with low spatial visualization in the virtual environment.

IV. RESULTS

The results reported in this paper are based on least squares regression models. The focus has been to look at two slightly different sets of variables. The first model looks at how a set of variables based on the environment and the participant impact performance measures (e.g., time and distance), while the second model adds a software flavor to the analysis by adding a variable that incorporates map resets and pans.

A. Regression based on Environment and Participants

The first model examines the impact of the environment (E = 0 Field or 1 Virtual), spatial ability (S = 0 High Visualization or 1 Low Visualization), and gender (G = 0 Female or 1 Male) on time and distance.

In particular we looked at $Y = E + S + E*S + G$, where Y is the prediction of either log(time) or log(distance) and E*S is the interaction of the environment and the participant's spatial ability assignment.

The regression results for log(time) and log(distance) are shown in Tables I and II, respectively.

TABLE I. REGRESSION RESULTS FOR LOG(TIME).

	Estimate	Std.Error	t-value	Pr(>|t|)
(Intercept)	3.338	0.112	29.814	0.000
Env	0.083	0.159	0.524	0.605
Spatial	0.398	0.152	2.622	0.014
Gender	0.005	0.107	0.050	0.961
Env:Spatial	0.014	0.209	0.068	0.946

TABLE II. REGRESSION RESULTS FOR LOG(DISTANCE).

	Estimate	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.021	0.117	0.177	0.861
Env	0.268	0.166	1.617	0.117
Spatial	0.373	0.159	2.355	0.026
Gender	0.029	0.111	0.261	0.796
Env:Spatial	-0.425	0.218	-1.944	0.062

The most interesting aspect of the results shown in the two tables is the significance of the participants' level of spatial ability in both results (Pr(>|t|) = 0.014 and 0.026, respectively. The regression model for distance (Table II) is suggestive that the interaction of E and S is important with Pr(>|t|) = 0.062.

The box plots for log(Time) and log(Distance) mediated by the interaction are shown in Figures 4 and 5, respectively. While the box plots don't provide too much information, they do provide some insight. First, in Figure 5 it appears that high spatial participants in the virtual environment traveled greater distances than the high spatial participants did in the field. The log(Time) boxplot (Figure 4) is suggestive that spatial ability is important in terms of the time taken (log(Time)). Such a result makes sense in light of the significance of spatial ability seen in Table I.

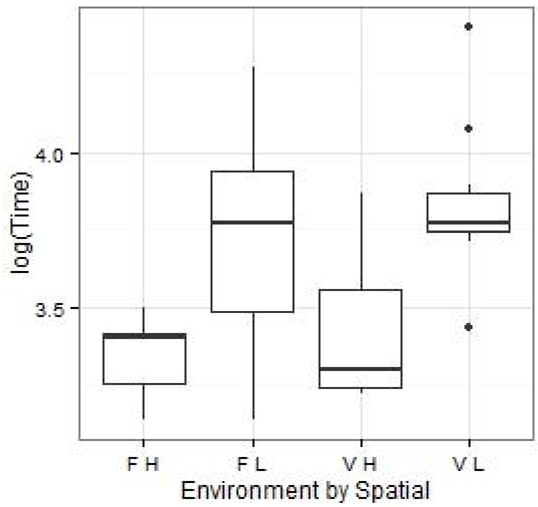

Figure 4. Boxplots for log(Time) vs E*S.

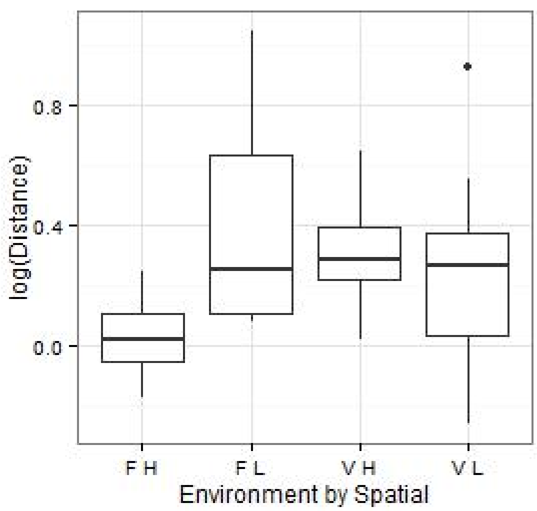

Figure 5. Boxplots for log(Distance) vs E*S.

B. Including Software Operations in the Regression

To bring the software performance of the participants into the regression model, we defined a variable (Resetpan) as a measure of how many times a participant interacted with the map on the handheld device. Formally, ResetPan is defined by

Resetpan=0.5(Resets-mean(Resets)/sd(Resets) + Pans-mean(Pans)/sd(Pans)).

The choice of map resets and pans was motivated by the experiment observers' experience in both the field and virtual environments. Tables III, IV, and V show the regression results for the least squares regression models for log(Time), log(Distance), and Errors, respectively.

TABLE III. REGRESSION RESULTS FOR LOG(TIME).

	Estimate	Std.Error	t-value	Pr(>|t|)
(Intercept)	3.111	0.12	25.979	0.000
Env	0.041	0.138	0.297	0.769
Spatial	0.301	0.134	2.235	0.034
Gender	0.039	0.093	0.424	0.675
Resetpans	0.036	0.011	3.213	0.003
Env:Spatial	-0.005	0.181	-0.026	0.979

TABLE IV. REGRESSION RESULTS FOR LOG(DISTANCE).

	Estimate	Std.Error	t-value	Pr(>|t|)
(Intercept)	-0.200	0.128	-1.563	0.130
Env	0.227	0.148	1.541	0.135
Spatial	0.278	0.144	1.932	0.064
Gender	0.062	0.099	0.626	0.537
Resetpans	0.035	0.012	2.915	0.007
Env:Spatial	-0.443	0.193	-2.292	0.030

TABLE V. REGRESSION RESULTS FOR ERRORS.

	Estimate	Std.Error	t-value	Pr(>|t|)
(Intercept)	2.219	0.772	2.875	0.008
Env	-0.123	0.888	-0.138	0.891
Spatial	1.811	0.867	2.089	0.047
Gender	0.338	0.597	0.565	0.577
Resetpans	-0.029	0.073	0.000	0.696
Env:Spatial	-0.919	1.164	-0.790	0.437

From the results in the three tables, it appears that the spatial ability levels of the participants was again an important variable as it is significant for log(Time) (Pr(>|t|) = 0.034 in Table III) and for Errors (Pr(>|t|) = 0.045 in Table V). It was also suggestive for log(Distance) (Pr(>|t|) = 0.064 in Table IV). As expected, Resetpans showed up as significant for both log(Time) and log(Distance) results. Interestingly, it doesn't show up as a factor in the number of errors made by the participants. Finally, as in Table II, the interaction between the environment and the level of spatial ability was significant for log(Distance) (Pr(>|t|) = 0.031 in Table IV).

V. DISCUSSION

The spatial visualization ability level of the participants was a significant factor in the regression models for all but the regression model for log(Distance) using the Resetpan variable and was suggestive there. The boxplots in Figures 4 and 5 show some support for the hypotheses given earlier except for the virtual environment participants in the log(Distance) boxplot. Somewhat surprising was how little impact the environment (and the interaction between the environment and spatial ability) had in the experiment. Beyond the suggestion of a difference in the distance traveled between high spatial visualization ability people in the field versus high spatial visualization ability people in the virtual environment (Figure 5), we did not find a separation based on the environment.

Two issues that made it difficult to work with the data were the number of participants used in the study and the high level of variation that we found in the performance of low spatial participants, especially in the virtual environment. The number of participants that we used in the experiment was a function of the difficulty that we had in finding participants with a low level of spatial ability. In spite of testing a large pool of subjects in Phase I, we struggled to find a sufficient number of low spatial ability participants to increase the size of the experiment.

VI. CONCLUSION AND FUTURE WORK

The experiments provided support for the development of map-based user interfaces that work with both high and low levels of spatial ability. Since we were unable to find a significant difference due to environment, testing software in the virtual environment remains a realistic possibility. However, we expect that we will have to do more testing with an increased Phase II sample size in order to validate such an approach.

VII. REFERENCES

[1] Andrews, E. C., & Westerman, S. J. (2012). Age differences in simulated driving performance: Compensatory processes. *Accident Analysis & Prevention*, *45*(0), 660–668.

[2] Arnold, P., & Farrell, M. J. (2003). Embodiment and spatial behavior in virtual environments: comments on Durlach et al.(2000). *Presence: Teleoperators & Virtual Environments*, *12*(6), 658–662.

[3] Benyon, D., Crerar, A., & Wilkinson, S. (2001). Individual differences and inclusive design. In C. Stephanidis (Ed.), *User interfaces for all: concepts, methods, and tools*. Mahwah, NJ: Lawrence Erlbaum Associates.

[4] Benyon, D. R. (1993). Accommodating individual differences through an adaptive user interface. In M. Schneider-Hufschmidt, T. Kuhme, & U. Malinowski (Eds.), *Adaptive User Interfaces: Principles and Practice* (pp. 149–165). Amsterdam: Elsevier Science.

[5] Carroll, J. (1993). *Human cognitive abilities: A survey of factor analytical studies.* Cambridge: Cambridge University Press.

[6] Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.

[7] Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: audio visual experience automatic virtual environment. *Commun. ACM*, *35*(6), 64–72. doi:10.1145/129888.129892

[8] Downing, R. E., Moore, J. L., & Brown, S. W. (2005). The effects and interaction of spatial visualization and domain expertise on information seeking. *Computers in Human Behavior*, *21*(2), 195–209.

[9] Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

[10] Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, *34*(2), 151–176.

[11] Jennings, F., Benyon, D. R., & Murray, D. M. (1991). Adapting systems to differences between individuals. *Acta Psychologica*, *78*(1-3), 243–258.

[12] Kozhevnikov, M., Motes, M., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology*, *20*, 397–417.

[13] Li, C. (2006). User preferences, information transactions and location-based services: A study of urban pedestrian wayfinding. *Computers, Environment and Urban Systems*, *30*(6), 726–740.

[14] Li, C., & Willis, K. (2006). Modeling context aware interaction for wayfinding using mobile devices. *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (pp. 97–100).

[15] Liu, A. M., Oman, C. M., Galvan, R., & Natapoff, A. (2012). Predicting space telerobotic operator training performance from human spatial ability assessment. *Acta Astronautica*, (0), -. doi:10.1016/j.actaastro.2012.04.004

[16] Lohman, D. F. (1996). Spatial ability and g. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 97–116). Mahwah, NJ: Lawrence Erlbaum Associates.

[17] Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods*, *31*(4), 557–564.

[18] Murray, H. A., & Kluckhohn, C. (1948). *Personality in nature, society, and culture*. Oxford, England: Knopf.

[19] Ni, T., Bowman, D. A., & Chen, J. (2006). Increased display size and resolution improve task performance in Information-Rich Virtual Environments. *Proceedings of Graphics Interface 2006* (pp. 139–146).

[20] Royce, J. R. (1973). The conceptual framework for a multi-factor theory of individuality. In J. R. Royce (Ed.), *Multivariate Analysis and Psychological Theory*. London and New York: Academic Press.

[21] Schellenbach, M., Lövdén, M., Verrel, J., Krüger, A., & Lindenberger, U. (2010). Adult age differences in familiarization to treadmill walking within virtual environments. *Gait & posture*, *31*(3), 295–299.

[22] Vicente, K. J., Hayes, B. C., & Williges, R. C. (1987). Assaying and isolating individual differences in searching a hierarchical file system. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *29*(3), 349–359.

[23] Witmer, B. G., & Kline, P. B. (1998). Judging perceived and traversed distance in virtual environments. *Presence*, *7*(2), 144–167.

[24] Wobbrock, J. O., Kane, S. K., Gajos, K. Z., Harada, S., & Froehlich, J. (2011). Ability-Based Design: Concept, Principles and Examples. *ACM Trans. Access. Comput.*, *3*(3), 9:1–9:27. doi:10.1145/1952383.1952384

[25] Zhang, H., & Salvendy, G. (2001). The implications of visualization ability and structure preview design for Web information search tasks. *International Journal of Human-Computer Interaction*, *13*(1), 75–95.

 ISBN: 978-1-61208-250-9