



Investigation of spatial ability test completion times in virtual reality using a desktop display and the Gear VR

Tibor Guzsvinecz¹ · Éva Orbán-Mihálykó² · Cecília Sik-Lányi¹ · Erika Perge³

Received: 28 February 2020 / Accepted: 20 February 2021
© The Author(s) 2021

Abstract

The interaction time of students who did spatial ability tests in a virtual reality environment is analyzed. The spatial ability test completion times of 240 and 61 students were measured. A desktop display as well as the Gear VR were used by the former group and by the latter one, respectively. Logistic regression analysis was used to investigate the relationship between the probability of correct answers and completion times, while linear regression was used to evaluate effects and interactions of following factors on test completion times: the users' gender and primary hand, test type and device used. The findings were that while the completion times are not significantly affected by the users' primary hand, other factors have significant effects on them: they are decreased by the male gender in itself, while they are increased by solving Mental Rotation Tests or by using the Gear VR. The largest significant increment in interaction time in virtual reality during spatial ability tests is when Mental Rotation Tests are accomplished by males with the Gear VR, while the largest significant decrease in interaction time is when Mental Cutting Tests are completed with a desktop display.

Keywords Cognitive skills · Desktop display · Gear VR · Human-computer interaction · Interaction time · Mental rotation · Spatial ability · Virtual reality

1 Introduction

A theory is proposed by Gardner (1983), saying that every human has multiple types of intelligence and spatial intelligence is one of them. This theory was improved by Maier (1996) who concluded that spatial intelligence is made up of five different parts: spatial perception, visualization, mental rotation, spatial relations and spatial rotations. According

to Miller and Bertoline (1991), this ability is not a biological susceptibility, but can be improved through time: improvement can occur simply by life experiences or by being exposed to certain learning environments. It has been suggested in the study of Miller (1992) that spatial ability training should be included in the curriculum of engineering studies. According to Ghiselli (1973), the success in the fields of engineering, mathematics and architecture is related to the spatial skills of the person.

A well-developed spatial ability is important in modern age and it becomes more relevant with each passing day as it is required by many jobs. With it, the person can understand relations between objects and space. A considerable amount of paper-based tests was developed through the years to improve spatial intelligence and ability of people. These tests include the Mental Rotation Test (MRT) (Ault and John 2010), the Mental Cutting Test (MCT) (Bosnyak and Nagy-Kondor 2008) and the Purdue Spatial Visualization Test (PSVT) (Branoff and Connolly 1999). Since these are paper-based tests, the following question can arise: “*what happens when these tests are taken in virtual reality? (VR)*”

This is an essential question as according to Burdea and Coiffet (2003), a VR system is made up of five important

✉ Tibor Guzsvinecz
guzsvinecz@virt.uni-pannon.hu

Éva Orbán-Mihálykó
orbane@almos.uni-pannon.hu

Cecília Sik-Lányi
lanyi@almos.uni-pannon.hu

Erika Perge
perge@eng.unideb.hu

¹ Department of Electrical Engineering and Information Systems, University of Pannonia, Veszprém, Hungary

² Department of Mathematics, University of Pannonia, Veszprém, Hungary

³ Department of Basic Technical Studies, University of Debrecen, Debrecen, Hungary

components: the VR engine itself, its software/database(s), I/O devices, tasks and users themselves. This means that users are as equally important as the other factors in a VR system (Heldal 2007; Schroeder et al. 2006). As users are part of this, interactions can occur between its components as shown by several studies: not only the learning skills of users can be increased due to VR (Horváth 2016; Kovari 2018; Wilson 2019), but their spatial skills as well (Dünser et al. 2006; Macik 2018; McLellan 1998; Molina-Carmona et al. 2018a; Parsons et al. 2004; Torner et al. 2016). The conclusion of the last two studies is that mental rotation of males are better than that of females in VR and in augmented reality (AR). Also, the suggestion of the last study is that AR could be a good tool for improving spatial ability.

Studies also exist which present the design of VR spatial ability tests and the research plan of their authors. An MCT test in VR was developed by Hartman et al. (2006) with the goal to help other scientists in the creation of future MCT tests. A testing method was outlined by presenting the procedure and data analysis. As the study is about the creation of VR MCT tests, their results are not published in this study. A VR MRT test outline was created by Rizzo et al. (1998b) by presenting their future plan and their first results at the time of writing the referenced paper: the rates of correct answers between a pre-test and a post-test were investigated in it. Later, their preliminary results on these tests were published (Rizzo et al. 1998a). According to them, the VR MRT test helped users as their results were improved on the post-tests. A study that has different results also exist in the literature (Jiang and Laidlaw 2019), where results of two groups which did the MRT test type were compared: a desktop environment was used by one group and a virtual environment (VE) by the other. According to them, low spatial ability participants benefited from learning between the pre-test and the post-test. The conclusion was that their performance on MRT tests were not significantly affected by using VR technologies.

However, in the study of Oman et al. (2000) it was found out that the performance of users who used head-mounted displays (HMDs) was slightly better than that of those who did not use them. Their conclusion was that VR can be used for spatial ability training and thus, is excellent for this purpose. According to Chang et al. (2017), a perspective-test was developed in VR to measure the spatial skills of users. Similarly to the previous study, pre-tests and post-tests were conducted. Users were grouped into three: those who interacted with the application with motion; those who interacted with a keyboard and a mouse; and users who interacted with motion, but used nonspatial tasks. Their conclusion was that the first two groups improved between the tests. However, significant improvements were only found in the case of the first group. The PSVT-R test in VR was created and evaluated by Molina et al. (2018b). Two groups of users

were tested: a desktop display was used by one group and an HMD by the other one. Pre-tests and post-tests were done by both of them. In the study, the conclusion was that there are improvements in the spatial skills of both groups, but it is significant when an HMD is used.

It is proven by our earlier research that results which are similar to those of paper-based tests can be gathered when using a desktop display (Guzsvinecz et al. 2020a). However, when using an HMD such as the Gear VR, they change significantly. This means that positive influence of an HMD is also confirmed, however new facts arose in the referenced study: while a significant difference exists between results of males and females when using a desktop display, this difference disappears when the Gear VR HMD is used. Moreover, a similar phenomenon can be observed between right-handed and left-handed users: in the case of a desktop display, the performance of right-handed people on the tests is significantly better than that of their left-handed counterparts, but with use of the Gear VR, significantly better results are achieved on tests by left-handed users than by right-handed ones.

As can be seen that while transition from paper-based tests to digital ones is not easy, it has certain advantages, especially on different user groups. This is because when a user is placed inside a VE, interaction between the human and machine changes: for example, the user does not use a pen and paper, but a sensor (or other input devices) to take tests. This also means that in VR the human-computer interaction (HCI) and human-computer interfaces can be different from application to application (Kortum 2008): new interfaces and various I/O devices have to be learned in every VR application and the required tasks could differ between each of them. Due to this, the developers of VR applications have to take these differences into account (Sutcliffe et al. 2019), and the focus should be on user-centric development (Drettakis et al. 2007). To help users with HCI, a toolkit was developed by Takala (2014) which makes it easier to create VR applications using building blocks. With it, applications can be created for HMDs. In this study, spatial graphical user interface ideas of students are presented and the toolkit is evaluated. According to them, both received positive feedback.

Therefore, to investigate multiple aspects of different types of user interaction in VR and influences of display devices and display parameters on it, a spatial ability measuring VR application was developed, which can use a desktop display and the Gear VR (Guzsvinecz et al. 2019). Regarding tests, since some factors and influences on correct answers were investigated in another paper (Guzsvinecz et al. 2020b), in this study the factors that affect the spatial ability test completion times are planned to be found. What can be extrapolated from the results is how to make interaction with the computer in VR less time consuming. As

mentioned in several studies (Chang et al. 2017; Guzsvinecz et al. 2020a; Molina et al. 2018b; Oman et al. 2000), using an HMD has positive influence on the users' ratio of correct answers, but this effect is different between user groups. Thus, it is possible that completion times could be affected by different display devices, users' various characteristics, and test types. If this is the case, designers of VR applications which require spatial skills (such as applications for education and even for cognitive rehabilitation) could use the results to create VEs which are less time-consuming and/or more user-friendly as users' characteristics are taken into account during development.

2 Research questions and hypotheses

As mentioned in the introductory section, complexity of a VR system can be testified to some extent by completion times and even rates of correct answers: such systems are very complex and users, tasks and I/O devices are integral components of them. Therefore, the test completion times in VR were investigated because using different display devices can have certain positive or negative influences on the results of correct answers of users with various characteristics. In case of this study, these human characteristics mean users' gender, and their primary hand. These two characteristics were chosen as significant differences were found in their rates of correct answers in our previously referenced research. In addition, the completion times of each test type are also investigated to see whether they interact with different display devices and various user groups.

Therefore, firstly, the connection between completion times and the probabilities of correct answers were investigated. Secondly, the influence of used display devices on completion times regarding various user groups and test types was needed to be analyzed. Lastly, the goal was to find the combinations of display devices, human characteristics, and certain test types which result in either the smallest or the largest completion times. Therefore, before the research commenced, three research questions (RQs) were set up, which are the following:

- RQ1: Are completion times and probabilities of correct answers independent from each other?
- RQ2: Are completion times affected by different display devices, users' various characteristics, and test types? Do they interact with each other?
- RQ3: Which combination of mentioned factors results in either the smallest or the largest test completion times?

After setting up the RQs, the same number of hypotheses (Hs) was formulated. These three Hs which contain both the null hypotheses and their alternatives, are the following:

- H1: Completion times and probabilities of correct answers are independent from each other, opposite to: completion times and probabilities of correct answers are dependent.
- H2: No significant effects on completion times are provided by different display devices, human characteristics and test types, opposite to: significant effects on completion times are provided by different display devices, human characteristics and test types.
- H3: The smallest and the largest completion times are not significantly affected by some of the mentioned factors, opposite to: the smallest and the largest completion times are significantly affected by some of the mentioned factors.

3 Methodology

In order to answer the mentioned RQs, a spatial ability testing application was developed at the University of Pannonia in 2019. The Unity game engine was used for development and two versions of spatial ability application exist: one was made for Windows 7 or newer operating systems and the other one was developed for Gear VR SM-R322 HMD, which uses Android operating system. A Samsung Galaxy S6 Edge+ smartphone was placed inside the Gear VR.

While the two versions of spatial ability testing applications are built similarly, two main differences exist between them. In the case of the desktop display version, interaction is done with a keyboard and/or mouse. When using the Gear VR however, interaction is done with a touchpad which can be found on its right side. As the I/O devices are integral components of a VR system, this is a critical difference. Another distinction is that the virtual camera can rotate when the Gear VR is used: the smartphone inside it has accelerometer(s) and gyroscope(s), thus rotation of users' head can be followed, meaning the virtual camera can rotate accordingly. Thus, all objects could be seen from slightly different perspectives and the students felt that they were inside the VE. However, the virtual camera cannot move in any direction due to the Gear VR being only able to handle rotations. In case of the desktop display version, the virtual camera could not be rotated or moved, thus the objects could only be seen from a frontal point of view: the immersion of users was not as high, because they were outside of the VE.

Testing was done in two groups during September 2019. At the University of Debrecen, an LG 20M37A (19.5") desktop display device was used for testing by 240 students. Those who tested at the University of Debrecen were either architect and civil engineering or mechanical engineering students. The ones who came to tests were 23.5 years old on average with a dispersion of 3.1 years. For all tests, a computer laboratory was used. Due to its small size, twelve

groups of twenty students were made. Testing was done during three weekdays and thus, was completed within a week. At the University of Pannonia, 61 users tested with the Gear VR: this group consisted of information technology (IT) and non-IT students. They were 19.7 years old on average with a dispersion of 1.5 years. In this case the tests' duration was three weeks long as only one Gear VR was available at the University. This means that all testers had to come in a sequential order, one-by-one. Each of them required at least thirty minutes and an hour at most to complete the tests. Thus, the skills of 8 students were measured at most per day, while the smallest number of testers per day was 2. As they were students, their appointments were made according to their classes so they could come to the tests before or after their classes.

During measurements, each test type had to be done three times – in other words, three sequences. Each test type could be found in every sequence. In all of them, the order was the MRT, MCT, PSVT test types. As each test type consisted of ten rounds, thirty questions could be found in every sequence. After one sequence of tests was completed, users could rest – if they wanted to – and after that they started the next round of testing which consisted of same test types, but their solutions to spatial ability problems were changed using a randomization technique. In total, each student was asked 90 questions on the tests. In Figs. 1, 2, 3 the MRT, MCT and PSVT test types can be seen in the application, respectively.

Naturally, on the tests, answers had to be chosen, while the completion times of users were also measured (as seen in the upper right corners in Figs. 1, 2, 3). This was done differently in each version of the application due to the two types of interactions. In case of the desktop display version, answers could be chosen by pressing certain numbers on the keyboard: 1–4 in case of the MRT test type and 1–5 during other test types. These numbers correspond to objects on screen from left to right. Selecting answers could also be done by clicking on a certain object with a mouse. However, when the Gear VR version is used, students had to look at

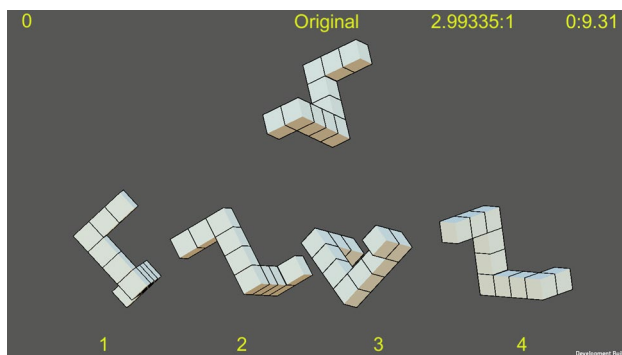


Fig. 1 Screenshot of the MRT test type

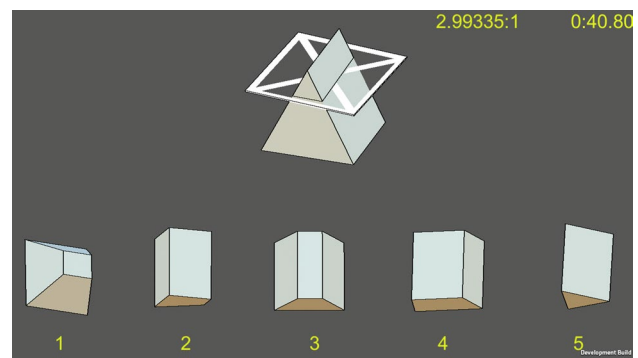


Fig. 2 Screenshot of the MCT test type

objects they wanted to choose. Afterward, the touchpad had to be tapped on the right side of the Gear VR to select an object.

As mentioned earlier, the virtual camera's position is locked in both versions, but it can be rotated when using the Gear VR. While the number of rotations was measured during testing, students were asked to not rotate the virtual camera to correctly investigate their spatial skills.

Regarding students at both universities, every person who was willing to do the tests could join. This means that there were no selection criteria applied. Moreover, since the spatial skills of students were measured, no information was gathered of their height and body weight. To respect their anonymity, their names were not gathered.

It should be noted that information is logged about users (age, gender, primary hand, years spent at a university, what do they study), the test (test type, completion time, number of correct answers), and used technical parameters (virtual camera type, its rotation, its field of view, the contrast ratio between foreground object and the background, whether shadows are turned on in the scene and the used device) into a .csv file. Since completion times are in our focus, therefore, the effects of users' gender, primary hand, the test type, and used device are investigated on them.

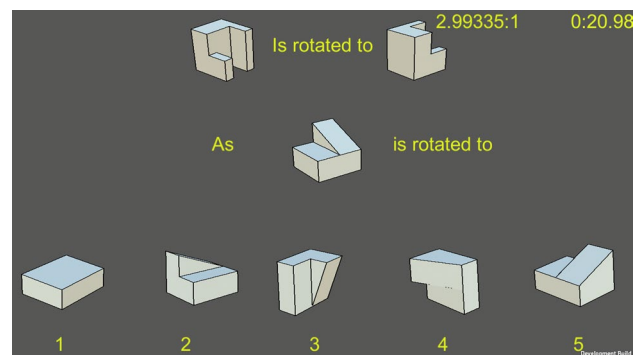


Fig. 3 Screenshot of the PSVT test type

As mentioned, the goal is to identify effects of the mentioned factors on test completion times. Thus, the correlation between probabilities of correct answers and test completion times is also investigated. To evaluate the relation between probabilities and completion times, logistic regression analysis method was used, while to analyze the factors' effects on the latter, linear regression analysis methods were used (Hosmer Jr et al. 2013; Walpole et al. 2011). All calculations were performed by help of statistical program package R (R Core Team 2018).

4 Results

In this section, completion times are analyzed in each case, and are measured in seconds. These are logged into the mentioned file after ten questions are answered (as one test type consists of ten questions). The smallest completion time is 7.9 seconds and the largest one is 1168.43 seconds which is approximately 20 minutes. Their average is 200.388 seconds with a dispersion of 123.279 seconds.

The distribution of completion times is not normal as the Kolmogorov-Smirnov test resulted in p -value $< 2.2 \times 10^{-16}$. Therefore, the hypothesis of normal distribution is rejected. The histogram of test completion times is presented in Fig. 4.

Since completion times and probabilities of correct answers are numerical values, the correlation coefficient can be used to evaluate whether they correlate or are independent.

The numerical value of correlation of completion times and probability of correct answers equals to 0.223. By

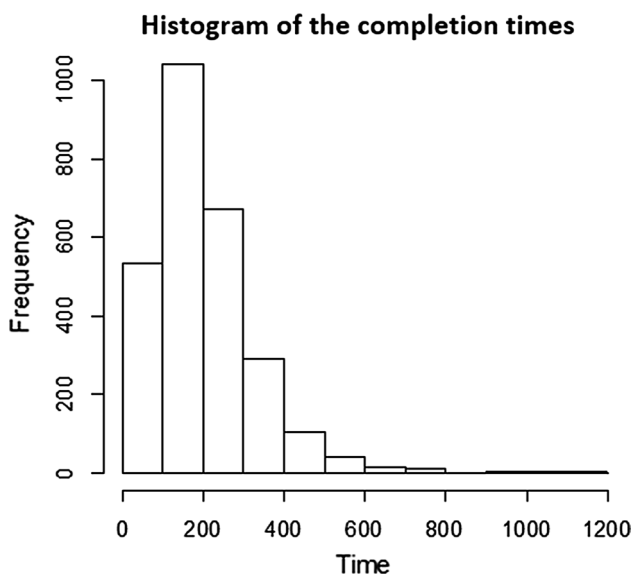


Fig. 4 The histogram of the spatial ability test completion times

performing a test to check whether it can be considered zero or not, p -value $< 2.2 \times 10^{-16}$ was received, therefore the hypothesis of correlation's zero value is rejected. This means that these variables are not independent of each other. The positive sign of correlation coefficient means that when completion time increases, the probability of correct answers increases as well. It is shown by the correlation coefficient's value that linear relationship is not strong between these two variables. If the correlation coefficient of the logarithm of completion time and probabilities is looked at, a somewhat larger correlation is yielded, which is 0.299.

Relations between completion times and the mentioned factors are analyzed by regression as seen in Table 1. In all Tables where the results of linear regression analysis are presented, the estimated coefficients (Est.), standard error (Std. err.), test statistics (t value) and p -value are shown in each case. It should be noted that the latter is the probability of type I. error ($\text{Pr}(> |t|)$). Inside those Tables where the results of logistic regression analysis are presented, z values are shown instead of t values.

According to Table 1, the influence of spatial ability test completion times is significant (p -value $< 2 \times 10^{-16}$). Meanwhile, according to the positive sign of estimated coefficient 0.002307, it can be safely stated that when the tests' time increases so does the probability of correct answers in tendency.

If the logarithm of spatial ability test completion times is analyzed by logistic regression, then the following coefficients are resulted as can be seen in Table 2:

The similarity of the two p -values is presented by Tables 1 and 2: both are less than 2×10^{-16} in case of completion times and their logarithm.

For detecting the effects of further parameters on time it took to finish tests, three different subsections are created: the effects of factors are investigated one-by-one in Sect. 4.1, while the influences of pairs and triplets are analyzed in

Table 1 Results of logistic regression analysis of the relation between completion times and probabilities of correct answers

	Est.	Std. err.	z value	$\text{Pr}(> z)$
Intercept	0.232039	0.022724	10.21	$< 2 \times 10^{-16}$
Time	0.002307	0.000099	3.19	$< 2 \times 10^{-16}$

Table 2 Results of the logistic regression analysis of the logarithm of the completion times

	Est.	Std. err.	z value	$\text{Pr}(> z)$
Intercept	-2.18981	0.08886	-24.64	$< 2 \times 10^{-16}$
$\log(\text{Time})$	0.56389	0.01726	32.68	$< 2 \times 10^{-16}$

Table 3 Numerical values of the completion times regarding the users' gender

Gender	Tests' number	Avg. time	Dispersion
Female (F)	414	223.644	123.214
Male (M)	2295	196.193	122.850

Table 4 Results of the regression analysis of the completion times by the gender of the users

	Est.	Std. err.	<i>t</i> value	Pr(> <i>t</i>)
Intercept	223.644	6.040	37.024	$< 2 \times 10^{-16}$
Male (M)	-27.451	6.563	-4.183	2.97×10^{-5}

Sects. 4.2 and 4.3, respectively. In the latter two subsections not only their effects, but their interactions are also traced.

4.1 Analyzing the effect of one factor

Since four factors exist in the investigation, this section is divided into four subsections. In them, the effects of users' gender, their primary hand, test type, and device used are investigated on completion times. These influences are evaluated one-by-one and are analyzed in Sects. 4.1.1, 4.1.2, 4.1.3 and 4.1.4, respectively.

4.1.1 The effect of users' gender on time

In this subsection, the effect of users' gender on time is investigated. See Table 3 for spatial ability test completion times of users regarding their gender.

As can be seen from the data presented in Table 3, the average completion times of female students are higher than that of male students. From this Table, it can be suspected that completion times are affected by the users' gender. To prove this suspicion, a regression analysis was done and its results are presented in Table 4.

According to Table 4, the users' gender has an effect which is significant (p -value = 2.97×10^{-5}). It is shown by the negative value of the coefficient belonging to males (-27.451) that their completion times are less than that of female ones. This latter fact can also be suspected from Table 3.

4.1.2 The effect of users' primary hand on time

The next to evaluate was the effect of users' primary hand on time. The time it took for users to complete every test is shown by Table 5, while the results of regression analysis can be seen in Table 6. In these two tables, left-handed users are abbreviated as LH, while right-handed ones are as RH.

Table 5 Numerical values of completion times regarding users' primary hand

Primary Hand	Tests' number	Avg. Time	Dispersion
LH	324	199.535	118.589
RH	2385	200.504	123.926

Table 6 Results of regression analysis of completion times by the primary hand of users

	Est.	Std. err.	<i>t</i> value	Pr(> <i>t</i>)
Intercept	199.5352	6.8501	29.129	$< 2 \times 10^{-16}$
RH	0.9685	7.3006	0.133	0.894

Table 7 Numerical values of completion times regarding test type

Test type	Tests' number	Avg. time	Dispersion
MCT	903	176.156	129.959
MRT	903	245.701	127.911
PSVT	903	179.307	96.374

Table 8 Results of regression analysis of completion times regarding test type

	Est.	Std. err.	<i>t</i> value	Pr(> <i>t</i>)
Intercept	176.156	3.963	44.454	$< 2 \times 10^{-16}$
MRT	69.545	5.604	12.410	$< 2 \times 10^{-16}$
PSVT	3.151	5.604	0.562	0.574

As can be seen in Table 5, numerical averages of both left-handed and right-handed users are almost the same. Therefore, it can be suspected that there is no significant relation between completion times and users' primary hand. After performing regression analysis, these expectations proved to be the case as can be seen in Table 6. Due to p -value = 0.894, there is no significant difference between times spent on tests.

Since this means that the completion times are not significantly affected by the users' primary hand, it is omitted from further analyses. Therefore, it will not be paired or grouped in triplets with other factors.

4.1.3 The effect of test type on time

In this subsection, the influence of test types on completion times is evaluated, which is shown in Tables 7 and 8. In the former, completion times regarding spatial ability test types are presented, while regression analysis' results are shown in the latter.

Table 9 Numerical values of completion times regarding the device used

Device used	Tests' number	Avg. time	Dispersion
DD	2160	188.784	121.389
GVR	549	246.043	120.117

Table 10 Results of the regression analysis of completion time by device used

	Est.	Std. err.	<i>t</i> value	Pr(> <i>t</i>)
Intercept	188.784	2.606	72.43	$< 2 \times 10^{-16}$
GVR	57.259	5.790	9.89	$< 2 \times 10^{-16}$

According to Table 7, the average time is numerically larger in case of MRT test type. After regression analysis was completed, it can be seen in Table 8 that the MRT test type's average completion time is significantly larger (p -value $< 2 \times 10^{-16}$) than of the others. When investigating average completion times of the MCT and PSVT test types, it can be observed that there is no detected significant difference between them. As can be seen from data presented in Table 7, these numerical values are almost equal.

4.1.4 The effect of device used on time

The last factor to be evaluated is the device used. In this subsection, its effect on completion times is evaluated. The numerical values of average time it took to finish tests and their dispersions are shown in Table 9, while regression analysis' results are presented in Table 10.

Based on average completion times that can be seen in Table 9, it is possible that completion times are significantly influenced by the device used. To prove this suspicion, a regression analysis was performed and its output is contained in Table 10.

According to Table 10, there is a detectable difference in test completion times between the two display devices: when using the Gear VR, a significant increase in completion times can be observed (p -value $< 2 \times 10^{-16}$). Therefore, the suspicion that arose due to the averages that are presented in Table 9 became true.

4.2 Analyzing the effect of pairs

In this subsection, influence of variable pairs is analyzed. Since users' primary hand was omitted from further analyses in Sect. 4.1.2 due to having no significant effects on completion times, only three pairs were made: pairs of gender and test type; of gender and device used; and of test type and

Table 11 Numerical results of users' completion times regarding their gender and test type

Pair	Tests' number	Avg. time	Dispersion
F-T1	138	200.152	135.476
F-T2	138	263.152	123.841
F-T3	138	207.628	98.435
M-T1	765	171.827	128.554
M-T2	765	242.553	128.458
M-T3	765	174.198	95.169

Table 12 Results of the regression analysis of users' gender and test type

	Est.	Std. err.	<i>t</i> value	Pr(> <i>t</i>)
Intercept	200.152	10.106	19.805	$< 2 \times 10^{-16}$
F-T2	63.000	14.292	4.408	1.08×10^{-5}
F-T3	7.476	14.292	0.523	0.600945
M-T1	-28.324	10.980	-2.580	0.009941
M-T2	42.401	10.980	3.862	0.000115
M-T3	-25.954	10.980	-2.364	0.018158

device used. These pairs are evaluated in Sects. 4.2.1, 4.2.2 and 4.2.3, respectively.

4.2.1 The effect of the pair of gender and test type on time

Out of the three variable pairs, all combinations of gender and test type were evaluated first. In Table 11 the numerical values of their completion times are presented. Starting from this point onward, genders and test types inside the tables are abbreviated: F stands for females and M for males, while T1 equals to the MCT, T2 to MRT and T3 to PSVT test types.

By looking at average completion times in Table 11, the pairs with the two smallest and the two largest completion times are the following: male users who did the MCT and PSVT test types have the smallest completion times, while males and females who accomplished the MRT test type have the largest ones. It can be suspected that some significant differences exist between completion times. To prove this suspicion, a regression analysis was performed. In Table 12 it is shown whether the time it took to finish these spatial ability tests is significantly influenced by these combinations.

Due to the results presented in Table 12, no significant differences can be detected between completion times of female users who did the MCT and PSVT tests. Meanwhile, there are significant differences in case of all other pairs: completion times of males who accomplished the MCT and PSVT tests are significantly smaller (p -value = 0.009941 and p -value = 0.018158, respectively), while they are

Table 13 Regression analysis results of completion time by pair of gender and test type, allowing interactions

	Est.	Std. err.	<i>t</i> value	Pr(> <i>t</i>)
Intercept	200.152	10.106	19.805	$< 2 \times 10^{-16}$
M	-28.324	10.980	-2.580	0.00994
T2	63.000	14.292	4.408	1.08×10^{-5}
T3	7.476	14.292	0.523	0.60095
M-T2	7.725	15.528	0.498	0.61885
M-T3	-5.106	15.528	-0.329	0.74231

Table 14 Numerical values of completion times regarding users' gender and device used

Pair	Tests' number	Avg. Time	Dispersion
F-DD	261	212.681	125.440
F-GVR	153	242.344	117.373
M-DD	1899	185.499	120.486
M-GVR	396	247.472	121.276

significantly larger in case of females and males who completed the MRT tests (p -value = 1.08×10^{-5} and p -value = 0.000115, respectively). This perfectly matches with the numerical values in Table 11 where this fact could be suspected earlier.

After comparing the linear regression model which contains only two factors to the one which takes their interactions into account, it can be concluded that these two models do not significantly differ from each other. The latter model is analyzed by regression and its results are presented in Table 13.

According to the results that are shown in Table 13, it can be concluded that users' gender and the MRT test type have influences. While they are significant, their interactions are not.

4.2.2 The effect of pair of gender and device used on time

The next combination of factors to analyze was users' gender and device used. Similarly as before, the numerical values of completion times and regression analysis' results can be seen in Tables 14 and 15, respectively.

Due to numerical values presented in Table 14, it can be suspected that significant differences exist between completion times of males who accomplished tests using a desktop display device and the other groups. When the Gear VR is used either by males or females, their average times it took to finish tests are unquestionably larger. Also, the average of test completion times in case of females who did tests with using a desktop display device is about halfway between those of the smallest and of the largest combinations.

Table 15 Results of the regression analysis of completion time by users' gender and device used

	Est.	Std. err.	<i>t</i> value	Pr(> <i>t</i>)
Intercept	212.681	7.484	28.417	$< 2 \times 10^{-16}$
F-GVR	29.663	12.312	2.409	0.016046
M-DD	-27.182	7.982	-3.405	0.000671
M-GVR	34.790	9.640	3.609	0.000313

Table 16 Regression analysis results by users' gender and device used, allowing interactions

	Est.	Std. err.	<i>t</i> value	Pr(> <i>t</i>)
Intercept	212.681	7.484	28.417	$< 2 \times 10^{-16}$
M	-27.182	7.982	-3.405	0.000671
GVR	29.663	12.312	2.409	0.016046
M-GVR	32.309	14.007	2.307	0.021150

According to the output of regression calculations, the completion times of both females and males who used the Gear VR are significantly (p -value = 0.016046 and p -value = 0.000313, respectively) increased, when compared to females who completed tests with a desktop display. However, it can also be significantly (p -value = 0.000671) decreased by the combination of males and a desktop display. These facts could also be suspected from the data presented in Table 14. The next two to compare were the average times of females and males who tested with the Gear VR. After the comparison was made, no significant differences were found between them as p -value = 0.3414.

Next, two various models were compared: only two factors were contained in the model (I), while the other took their interactions into account as well (II). Variance analysis resulted in p -value = 0.02107, meaning that the latter is more appropriate. Regression analysis was performed, and its output can be seen in Table 16.

According to the data shown in Table 16, both factors have significant effects on completion times. Besides the seen influences, interaction between them is also detectable. This coincides with the statement that the phenomenon is described better in the model in which interactions are allowed than in the one where they are not. Moreover, the value of estimated coefficient is 32.309. According to this, the interaction of gender and device used pair is very strong. Therefore, it can be safely stated that when tests are taken by male students with the Gear VR, the increase in completion times is much more than in case of females who used the same display device.

Table 17 Numerical values of users' completion times regarding test type and device used

Pair	Tests' number	Avg. Time	Dispersion
T1-DD	720	160.512	122.476
T1-GVR	183	237.708	140.238
T2-DD	720	235.983	130.113
T2-GVR	183	283.937	111.186
T3-DD	720	169.858	94.485
T3-GVR	183	216.483	94.980

Table 18 Results of the regression analysis of test type and device used

	Est.	Std. err.	t value	Pr(> t)
Intercept	160.512	4.351	36.889	$< 2 \times 10^{-16}$
T1-GVR	77.197	9.666	7.987	2.03×10^{-15}
T2-DD	75.471	6.154	12.265	$< 2 \times 10^{-16}$
T2-GVR	123.425	9.666	12.769	$< 2 \times 10^{-16}$
T3-DD	9.346	6.154	1.519	0.129
T3-GVR	55.971	9.666	5.791	7.82×10^{-9}

4.2.3 The effect of the pair of test type and device used on time

The last combination of pairs to analyze was the test type and device used. The numerical results and of the regression analysis are shown in Tables 17 and 18, respectively.

What can be seen from the data presented in Table 17 is that using a desktop display device results in smaller averages of completion times – except in case of the MCT test type. The largest average time to finish spatial ability assessments can be achieved on the MRT test type using the Gear VR. Due to the data presented in Table 17, it can be suspected that significant differences exist.

When comparing to the pair of MCT test type and desktop display (which is named as Intercept variable in this case), it can be concluded that every pair – except combination of the PSVT test type and use of the desktop display (T3-DD) – significantly affects completion times.

Afterward, two linear models were compared to each other: the first one only consists of the factors themselves (I) and in the second their interactions are taken into account as well (II). The latter proved to be superior with p -value = 0.04093. Therefore, it is presented in detail inside Table 19.

It is shown by regression analysis in Table 19 that both factors have influences and even interactions exist between them. The only one exception is when the PSVT test type and use of a desktop display device are combined. It is presented by the negative sign of estimated coefficient -29.243

Table 19 Regression analysis results of completion times by the pair of test type and device used, allowing interactions

	Est.	Std. err.	t value	Pr(> t)
Intercept	160.512	4.351	36.889	$< 2 \times 10^{-16}$
T2	75.471	6.154	12.265	$< 2 \times 10^{-16}$
T3	9.346	6.154	1.519	0.1289
GVR	77.197	9.666	7.897	2.03×10^{-15}
T2-GVR	-29.243	13.669	-2.139	0.0325

Table 20 Numerical values of completion times of users regarding all factors

Triplet	Tests' number	Avg. time	Dispersion
F-T1-DD	87	183.741	135.609
F-T1-GVR	51	228.147	131.875
F-T2-DD	87	254.074	128.061
F-T2-GVR	51	278.636	115.875
F-T3-DD	87	200.229	100.221
F-T3-GVR	51	220.250	94.948
M-T1-DD	633	157.319	120.327
M-T1-GVR	132	241.403	143.653
M-T2-DD	633	233.496	130.296
M-T2-GVR	132	285.985	109.705
M-T3-DD	633	165.683	92.979
M-T3-GVR	132	215.028	95.313

that the increase in completion times is less in case of MRT test type and Gear VR pair, than it would be expected by the sum of separate effects of this pair of variables.

4.3 Analyzing the effect of all factors on time

In this subsection, the effects of all factors are analyzed, which are grouped into a triplet. Since the users' primary hand was omitted from further analyses due to having no significant influences, only one combination could be created: users' gender, test type, and device used. The numerical values of users' completion times are shown in Table 20, while the output of regression analysis is presented in Table 21.

It can be seen in Table 20 that averages of completion times are less than 200 seconds in case of males and females who accomplished the MCT type of tests when they used a desktop display. The time it took for males to finish the PSVT test type with the same display device is also below 200 seconds, while it is a little more than that number in case of females. The averages are greatly above 200 seconds in the other groups. It can also be suspected from this table that significant differences exist between completion times.

Comparisons were done to F-T1-DD which was the Intercept variable in this case. The results are the following:

Table 21 Results of the regression analysis by all factors

	Est.	Std. err.	<i>t</i> value	Pr(> <i>t</i>)
Intercept	183.74	12.50	14.700	$< 2 \times 10^{-16}$
F-T1-GVR	44.41	20.56	2.160	0.030886
F-T2-DD	70.33	17.68	3.979	7.11×10^{-5}
F-T2-GVR	94.90	20.56	4.615	4.11×10^{-6}
F-T3-DD	16.49	17.68	0.933	0.351047
F-T3-GVR	36.51	20.56	1.776	0.075911
M-T1-DD	-26.42	13.33	-1.982	0.047586
M-T1-GVR	57.66	16.10	3.581	0.000348
M-T2-DD	49.76	13.33	3.732	0.000194
M-T2-GVR	102.24	16.10	6.350	2.51×10^{-10}
M-T3-DD	-18.06	13.33	-1.355	0.175688
M-T3-GVR	31.29	16.10	1.943	0.052091

completion times are significantly smaller in case of males who did MCT tests with a desktop display (M-T1-DD, p -value = 0.047586), while they are significantly larger in cases of females who completed MCT tests using the Gear VR (F-T1-GVR, p -value = 0.030886), females who finished MRT tests with either a desktop display (F-T2-DD, p -value = 7.11×10^{-5}) or the Gear VR (F-T2-GVR, p -value = 4.11×10^{-6}), males who accomplished MCT tests with the Gear VR (M-T1-GVR, p -value = 0.000348) and males who did MRT tests using either a desktop display (M-T2-DD, p -value = 0.000194) or the Gear VR (M-T2-GVR, p -value = 2.51×10^{-10}). The largest increase in completion times is in case of males who finished MRT tests with the Gear VR (M-T2-GVR). This largest increase is 102.24 seconds, while the largest decrease is 236.42 seconds (M-T1-DD). Therefore, due to regression analysis, the suspicion from Table 20 proved to be true.

Afterward, comparison of three different models began: the one in which only the factors' influences can be found (I), the other where interactions of pairs are allowed (II) and the third in which interactions of combination of triplets are permitted (III). After the comparison was completed, the conclusion is that out of models I and II, the latter is significantly better with p -value = 0.0069. However, the difference between models II and III is not significant (p -value = 0.9043). Therefore, model II is shown in detail inside Table 22.

It is shown by the linear regression analysis method that out of genders, completion times are significantly decreased in case of male users (p -value = 0.000417). Also, they are significantly increased by the MRT test type (p -value $< 2 \times 10^{-16}$). A similar phenomenon is observable when the Gear VR is used (p -value = 0.000504).

It can be concluded again that in case of male users, the increase in completion times is larger if the Gear VR is used than in case of females (p -value = 0.016756). Moreover,

Table 22 Regression analysis results of all variables, allowing interactions

	Est.	Std. err.	<i>t</i> value	Pr(> <i>t</i>)
Intercept	184.409	8.037	22.944	$< 2 \times 10^{-16}$
M	-27.182	7.692	-3.534	0.000417
T2	75.471	6.141	12.289	$< 2 \times 10^{-16}$
T3	9.346	6.141	1.522	0.128172
GVR	49.601	14.241	3.483	0.000504
M-GVR	32.309	13.499	2.394	0.016756
T2-GVR	-29.243	13.642	-2.144	0.032161
T3-GVR	-30.572	13.642	-2.241	0.025112

when using this device, this increase is smaller in case of the MRT and PSVT test types compared to MCT tests. Both of their interactions are significant with p -value = 0.032161 and p -value = 0.025112, respectively.

5 Discussion

According to the results presented in previous section, two hypotheses are rejected and one is a mixed case. While no accepted hypotheses exist based on the observations, it should be noted that accepting a basis hypothesis means that there is no effect and its rejection affirms the effect. In case of this study, the ones that are rejected are H1 and H3, while a mixed case is presented by H2. Naturally, the importance of results is also assessed in this section.

Therefore, this section is divided into three subsections. In Sect. 5.1 the rejected hypotheses can be seen as those are the ones where significant effects are detected, the mixed case is presented in Sect. 5.2, while the importance of our results is shown in Sect. 5.3.

5.1 Rejected hypotheses - detected effects

The first hypothesis to be rejected is H1 which talks about the probabilities of correct answers and test completion times being independent of each other. This was rejected at the beginning of Sect. 4, where a correlation test was done. Since it resulted in 0.223 with p -value $< 2.2 \times 10^{-16}$, it can be concluded that these two quantities are not independent of each other. The same result (dependency) was concluded by logistic regression. Therefore, T1 is formed: probabilities of correct answers and completion times are not independent of each other as p -value $< 2.2 \times 10^{-16}$.

The second rejected hypothesis is H3, which hypothesized that the smallest and the largest completion times are not significantly influenced by some of the mentioned factors. From Tables 3 and 4, it can be concluded that a significantly smaller average of completion times is produced

by males than it is by females. Contrarily, it is proven by Tables 5 and 6 that completion times are not significantly affected by users' primary hand, therefore this variable can be omitted. However, the fact that they are significantly increased by MRT tests and the Gear VR, is proved to be true in Tables 7, 8, 9, 10.

Even though these times are decreased by the male gender in itself, the fact is that when this factor is paired with MRT tests (Table 12), the time it took to finish them is actually increased. When grouped into a triplet with the Gear VR and MRT test type in Table 21, it has the largest increase in completion times as well. To know the smallest completion times, the factors that decrease them are needed to be investigated: in itself, they are only decreased by the male gender (Table 4). When paired with the MCT or PSVT test types, they are significantly decreased with p -value = 0.009941 and p -value = 0.018158, respectively. If paired with a desktop display (Table 15), they are significantly (p -value = 0.000671) decreased again. To get the final result, it has to be grouped into a triplet. Therefore, from Table 21, it can be concluded that when doing MCT and PSVT test types using a desktop display, they also decrease. However, only the triplet with MCT tests has a significant (p -value = 0.047586) decrease in completion times. Furthermore, this decrease is larger than in case of the triplet with PSVT tests. Out of these facts, T3 is formed: the spatial ability test completion times are significantly (p -value = 2.51×10^{-10}) increased by the combination of male gender, MRT test type and use of Gear VR which is also the largest increment, while they are significantly (p -value = 0.047586) decreased by the combination of male gender, MCT test type and use of a desktop display which is also the largest decrease.

5.2 The mixed case

A mixed case is presented by only one hypothesis, which is H2. First, the part about gender not having an influence was rejected due to Tables 3 and 4. In the former, average completion times suggested that gender has an effect, as the average completion time of female users is 223.644 seconds, while the average completion time of male users is 196.193 seconds. The difference between average times is verified by the latter table. According to the coefficient -27.451 in case of male users, they have smaller test completion times than females. Additionally, due to p -value = 2.97×10^{-5} , the difference is significant. Consequently, the effect of gender on test completion times is significant.

Next, primary hand's effect was assessed. After the regression analysis, it is shown in the second block of Table 6 (although it could be suspected from Table 5) that completion times are not significantly (p -value = 0.894) affected by the users' primary hand. This means that this

part of the hypothesis was accepted, hence it became a mixed case.

Afterward, the influence of test type was investigated. From Table 7, it could already be suspected that completion times are affected by the test types, as MRT tests have the largest average of time. There is a 39.479% increase in time between MRT and MCT test types, a 37.028% increase between MRT and PSVT test types and an 1.788% between MCT and PSVT test types. After the regression analysis, it is proven by the findings in Table 8 that the completion times are significantly (p -value < 2×10^{-16}) increased by MRT tests. Therefore, this part of the hypothesis was rejected.

Lastly, the used display device's effect was analyzed. Similarly to the previous cases, the fact that this part of the hypothesis can be rejected may also be suspected from numerical values that are presented in Table 9. Users' average test completion times are larger in case of the Gear VR than when using a desktop display. The difference between them is 30.330% which is quite high. The regression analysis in Table 10 proves the suspicion of significance with p -value < 2×10^{-16} .

Therefore, based on the analyses, T2 can be formed: completion times are significantly affected by the users' gender (p -value = 2.97×10^{-5}), test type (p -value < 2×10^{-16}) and used display device (p -value < 2×10^{-16}), while they are not significantly influenced by users' primary hand (p -value = 0.894).

5.3 The importance of the results

As mentioned in the introductory section, a VR system is made up of five parts: VR engine, its software/database(s), I/O devices, tasks, and the users themselves (Burdea and Coiffet 2003). This means that depending on tasks (which are the test types in this case), various groups of users can interact differently, slower or faster with the I/O devices (and thus, with the VR application itself). Since the spatial skills of users can be improved by VR (Chang et al. 2017; Dünser et al. 2006; Molina-Carmona et al. 2018a), it is important to note that various display devices can interact with users by having positive or negative effects on their test results. Clearly, the influence of an HMD is significantly greater than of a desktop display (Molina-Carmona et al. 2018b; Guzsvinecz et al. 2020b), and with the former type of display device the performance of users are enhanced on spatial ability tests (Oman et al. 2000). The display device's effect on users also depends on their various characteristics (Guzsvinecz et al. 2020a). Also, the correct answers can be influenced by the users' gender itself (Parsons et al. 2004), and the completion times are also affected by it as could be observed in the previous section. This means that the users themselves are important (Heldal 2007; Schroeder et al.

2006). As they are part of a complex system, it is natural that there is an interaction among all factors.

Although the number of virtual camera rotations was also measured due to the nature of the Gear VR version, only those results were investigated where it is ≤ 3 (as students were asked to not rotate the virtual camera). This value was chosen due to possible rotations by mistake, however in $\approx 85\%$ of the measured data it was 0. It is also important to note that while an improvement exists in the results when the Gear VR is used, there is no significant difference between them on the desktop display and Gear VR versions (Guzvinecz et al. 2020a).

However, it became apparent that the rates of correct answers of users are not the only ones that are affected, but their completion times as well. These are significantly (p -value = 2.97×10^{-5}) influenced by users' gender, making the completion times of males smaller by 12.274% than that of females. They are also significantly (p -value $< 2 \times 10^{-16}$) affected by the MRT test type, meaning that they are increased by an increment of 39.479% and 37.028%, when comparing it to MCT and PSVT test types, respectively. Lastly, a similar case can be seen among display devices: compared to a desktop display, spatial ability test completion times are significantly (p -value $< 2 \times 10^{-16}$) increased by 26.3365% when using the Gear VR.

Also, completion times and probabilities of correct answers are not independent of each other as mentioned in Sect. 5.1. This means that if users are given much time for test completion, their probability of answering correctly becomes higher. However, since completion times are increased using Gear VR as it needs to be interacted with differently than a traditional keyboard and mouse, more time on tests has to be given. This means that the test examination committee has to give enough, but not too much time for testers. Due to this, the following future research possibility is also presented: *which is the optimal test deadline?*

The fact that VR has a future in education is shown by most studies that are referenced in the introductory section as the use of an HMD can enhance the users' spatial skills. According to all observations, it can be safely stated that time spent on tests is the other side of the same coin: both the rates of correct answers and completion times are equally important while analyzing results, however due to complexity of a VR system, the latter also has to be taken into account in all cases. Users' gender, test type and display device influence the completion times, therefore spatial ability test deadlines also have to be adjusted accordingly when they are done in a VR system.

Nowadays, paper-based versions of these tests are also part of the curriculum of engineering studies. However, these tests can also be taken in a VE, but significant differences in completion times have to be noted due to complexity of a VR system. Therefore, when engineering education

and/or training for a job that require well-developed spatial skills are conducted in a VE, the results presented in this study can help to choose deadlines for tests. This is because the average completion times in a VE can be influenced by users' gender, test types and display devices.

As mentioned earlier, only those results were evaluated where the number of rotations is ≤ 3 to correctly assess spatial skills of users. Since the number of rotations of the virtual camera was also measured and more measurements are still in progress, the physical spatial skills of students could be investigated as well in the future. It would be interesting to see whether a difference exists between the results.

6 Conclusions

The completion time regarding spatial ability tests is complex and a lot of information can be hidden in it. Due to this fact, it is not affected by one investigated factor only. Therefore, evaluations were conducted one-by-one, then in pairs, and lastly, in a triplet.

From the one-by-one analyses, it was found out that the completion times of males are significantly less than that of females. Additionally, this is significantly increased by answering spatial ability problems on MRT tests or simply by using the Gear VR.

While improvements on users' skills are featured by the Gear VR (based on previous studies), the time it took for users to complete these spatial ability measuring tests is also increased with its use. This means that interaction is less time-consuming with the traditional keyboard and mouse than with touchpad on the right side of the Gear VR. However, it is possible that this interaction time can still be decreased in the future with new input devices.

Even though interaction with the Gear VR HMD is more time consuming than with a desktop display device, it should be noted that the increase in completion times between test types is less when an HMD is used. However, no VR system exists with only one factor: in a classical model, it is made up of five components. As humans and tasks are important parts of a VR system, their influence on completion times should be investigated as well. In this case the former consists of the users' gender and primary hand, while the latter equals to test type. Additionally, display devices were investigated as they are part of a classical VR system as well. When adding these factors together, the conclusion is the following:

The largest increase in interaction time in VR during the spatial ability tests is when males do MRT tests using the Gear VR, while the largest decrease in interaction time is when males do the MCT test type with a desktop display. These are the factors that have significant effects on spatial ability test completion times. It also has to be kept in mind

that probabilities of correct answers and completion times are not independent of each other: the former can be higher when the latter is larger.

Based on the findings presented, the conclusion is as follows: interaction time of males in VR environments is increased in a larger degree than in case of females when the Gear VR is used. Moreover, when it is used instead of a desktop display, similar values among the completion times of males and females can be observed.

Thus, these times on spatial ability tests were investigated. Since rotation of the virtual camera is also measured and more measurements are still in progress, the physical spatial skills of students can be assessed in the future as well.

Acknowledgements The authors acknowledge the financial support of Széchenyi 2020 under the EFOP-3.6.1-16-2016-00015. Project TKP2020-NKA-10 has been implemented with support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2020-4.1.1-TKP2020 Thematic Excellence Programme 2020 – National Challenges sub-program funding scheme. The authors would like to thank this support. The authors would also like to thank Ms. Mónika Szeles for her help in developing the application as well as Mr. Lóránt Horváth for his help in creation of objects for MCT test type. Last, but not least, the authors would like to thank everyone who helped by testing the spatial ability application.

Funding Open access funding provided by University of Pannonia.

Declarations

Conflict of interest The authors declare that they have no conflict of interest. The participants in tests gave their consent to gather information about their gender, age, primary hand, what do they study and how long they have attended a university at the time of tests. To respect their anonymity, their names were not gathered.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ault HK, John S (2010) Assessing and enhancing visualization skills of engineering students in africa: a comparative study. *Eng Des Graph J* 74:12–20
- Bosnyak A, Nagy-Kondor R (2008) The spatial ability and spatial geometrical knowledge of university students majored in mathematics. *Acta Didact Univ Comenianae Math* 8:1–25
- Branoff TJ, Connolly E (1999) The addition of coordinate axes to the purdue spatial visualization test-visualization of rotations: A study at two universities. In: *Proceedings of ASEE annual conference & exposition*, pp 1–9
- Burdea GC, Coiffet P (2003) *Virtual reality technology*. John Wiley, New Jersey
- Chang JSK, Yeboah G, Doucette A, Clifton P, Nitsche M, Welsh T, Mazalek A (2017) Evaluating the effect of tangible virtual reality on spatial perspective taking ability. In: *Proceedings of the 5th symposium on spatial user interaction*, pp 68–77. <https://doi.org/10.1145/3131277.3132171>
- Drettakis G, Roussou M, Reche A, Tsingos N (2007) Design and evaluation of a real-world virtual environment for architecture and urban planning. *Presence Teleoperators Virtual Environ* 16:318–332. <https://doi.org/10.1162/pres.16.3.318>
- Dünser A, Steinbügl K, Kaufmann H, Glück J (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*, pp 125–132. <https://doi.org/10.1145/1152760.1152776>
- Gardner H (1983) *Frames of mind: the theory of multiple intelligences*. Basic Books, New York
- Ghiselli EE (1973) The validity of aptitude tests in personnel selection. *Pers Psychol* 26:461–477. <https://doi.org/10.1111/j.1744-6570.1973.tb01150.x>
- Guzsvinecz T, Szeles M, Perge E, Sik-Lanyi C (2019) Preparing spatial ability tests in a virtual reality application. In: *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pp 363–368. <https://doi.org/10.1109/CogInfoCom47531.2019.9089919>
- Guzsvinecz T, Orban-Mihalyko E, Perge E, Sik-Lanyi C (2020a) Analyzing the spatial skills of university students with a virtual reality application using a desktop display and the gear vr. *Acta Polytech Hungarica*, 17:35–56. <https://doi.org/10.12700/APH.17.2.2020.2.3>
- Guzsvinecz T, Sik-Lanyi C, Orban-Mihalyko E, Perge E (2020b) The influence of display parameters and display devices over spatial ability test answers in virtual reality environments. *Appl Sci* 10:526–552. <https://doi.org/10.3390/app10020526>
- Hartman NW, Connolly PE, Gilger JW, Bertoline GR, Heisler J (2006) Virtual reality-based spatial skills assessment and its role in computer graphics education. In: *ACM SIGGRAPH 2006 Educators program*, pp 46–53. <https://doi.org/10.1145/1179295.1179342>
- Heldal I (2007) The impact of social interaction on usability for distributed virtual environments. *Int J Virtual Real* 6(3):45–54
- Horváth I (2016) Innovative engineering education in the cooperative vr environment. In: *2016 7th IEEE international conference on cognitive infocommunications (CogInfoCom)*, pp 359–364. <https://doi.org/10.1109/CogInfoCom.2016.7804576>
- Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) *Applied logistic regression*, vol 398. John Wiley, New Jersey
- Jiang E, Laidlaw DH (2019) Practicing in virtual reality improves mental rotation ability: lower scorers benefit more. *Brown University*. https://cs.brown.edu/research/pubs/theses/ugrad/2019/jiang_elaine.pdf. Accessed 12 October 2020
- Kortum P (2008) *HCI beyond the GUI: design for haptic, speech, olfactory, and other nontraditional interfaces*. Elsevier, Amsterdam
- Kovari A (2018) Coginfocom supported education: a review of coginfocom based conference papers. In: *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pp 233–236. <https://doi.org/10.1109/COGINFOCOM.2018.8639879>
- Macik M (2018) Cognitive aspects of spatial orientation. *Acta Polytech Hungarica* 15:149–167. <https://doi.org/10.12700/APH.15.5.2018.5.9>

- Maier PH (1996) Spatial geometry and spatial ability—how to make solid geometry solid. In: Selected papers from the annual conference of didactics of mathematics, pp 63–75
- McLellan H (1998) Cognitive issues in virtual reality. *J Vis Lit* 18:175–199. <https://doi.org/10.1080/23796529.1998.11674538>
- Miller CL (1992) Enhancing visual literacy of engineering students through the use of real and computer generated models. *Eng Des Graph J* 56:27–38
- Miller CL, Bertoline GR (1991) Spatial visualization research and theories: their importance in the development of an engineering and technical design graphics curriculum model. *Eng Des Graph J* 55:5–14
- Molina-Carmona R, Pertegal-Felices ML, Jimeno-Morenilla A, Mora-Mora H (2018a) Assessing the impact of virtual reality on engineering students' spatial ability. In: The future of innovation and technology in education: policies and practices for teaching and learning excellence, pp 171–185. <https://doi.org/10.1108/978-1-78756-555-520181013>
- Molina-Carmona R, Pertegal-Felices ML, Jimeno-Morenilla A, Mora-Mora H (2018b) Virtual reality learning activities for multimedia students to enhance spatial ability. *Sustain* 10:1074–1086. <https://doi.org/10.3390/su10041074>
- Oman CM, Shebilske WL, Richards JT, Tubré TC, Beall AC, Natapoff A (2000) Three dimensional spatial memory and learning in real and virtual environments. *Spat Cogn Comput* 2:355–372. <https://doi.org/10.1023/A:1015548105563>
- Parsons TD, Larson P, Kratz K, Thiébaux M, Bluestein B, Buckwalter JG, Rizzo AA (2004) Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia* 42:555–562. <https://doi.org/10.1016/j.neuropsychologia.2003.08.014>
- R Core Team (2018) R: A language and environment for statistical computing. Vienna, Austria. <http://cran.univ-paris1.fr/web/packages/dplR/vignettes/intro-dplR.pdf>. Accessed 14 October 2020
- Rizzo A, Buckwalter JG, Larson P, Van Rooyen A, Kratz K, Neumann U, Kesselman C, Thiebaux M (1998a) Preliminary findings on a virtual environment targeting human mental rotation/spatial abilities. In: Proceedings of the 2nd european conference on disability, virtual reality and associated techniques, Sköve, Sweden, pp 213–220
- Rizzo AA, Buckwalter JG, Neumann U, Kesselman C, Thiébaux M, Larson P, van Rooyen A (1998b) The virtual reality mental rotation spatial skills project. *Cyberpsychology Behav* 1:113–119. <https://doi.org/10.1089/cpb.1998.1.113>
- Schroeder R, Heldal I, Tromp J (2006) The usability of collaborative virtual environments and methods for the analysis of interaction. *Presence Teleoperators Virtual Environ* 15(6):655–667. <https://doi.org/10.1162/pres.15.6.655>
- Sutcliffe AG, Poullis C, Gregoriades A, Katsouri I, Tzanavari A, Herakleous K (2019) Reflecting on the design process for virtual reality applications. *Int J Hum Comput Interact* 35:168–179. <https://doi.org/10.1080/10447318.2018.1443898>
- Takala TM (2014) Ruis: A toolkit for developing virtual reality applications with spatial interaction. In: Proceedings of the 2nd ACM symposium on Spatial user interaction, pp 94–103. <https://doi.org/10.1145/2659766.2659774>
- Torner J, Alpiste F, Brigos M (2016) Virtual reality application to improve spatial ability of engineering students. In: WSCG'2016-24th International conference in central europe on computer graphics, visualization and computer vision'2016, pp 69–77
- Walpole RE, Myers RH, Myers SL, Ye K (2011) Probability & statistics for engineers & scientists. Pearson Prentice Hall, New Jersey
- Wilson A (2019) Analysis of current virtual reality methods to enhance learning in education. *Sel Comput Res Pap* 8:61–66

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.