Analyzing the Spatial Skills of University Students with a Virtual Reality Application using a Desktop Display and the Gear VR

Tibor Guzsvinecz[1], Éva Orbán-Mihálykó[2], Erika Perge[3], Cecilia Sik-Lányi[4]

[1,4]Department of Electrical Engineering and Information Systems, Faculty of Information Technology, University of Pannonia, Egyetem utca 10, 8200 Veszprém, Hungary

[2]Department of Mathematics, Faculty of Information Technology, University of Pannonia, Egyetem utca 10, 8200 Veszprém, Hungary

[3]Department of Basic Technical Studies, Faculty of Engineering, University of Debrecen, Ótemető utca 2, 4028 Debrecen, Hungary

[1]guzsvinecz@virt.uni-pannon.hu, [2]orbane@almos.uni-pannon.hu, [3]perge@eng.unideb.hu, [4]lanyi@almos.uni-pannon.hu

Abstract: This paper presents the analyzed results of a virtual reality application which allows measuring spatial skills of the users. The application contains mental rotation tests of three types and can be used on a computer with desktop displays and on Android with the Gear VR device. The authors measured the spatial ability of 61 students with the Gear VR and of 240 students with a desktop display. By investigating the correct answer ratios, comparisons were done between the age, the gender, the primary hand, what the students are majoring in, moreover the display devices. The use of the Gear VR significantly improved on the average performance of female students by 18.02%, on left-handed students by 18.66%, on older students by 7.29% and made the purdue spatial visualization test easier by 17.21% compared to the use of a desktop display. These results also strengthen the fact that education has a future in virtual reality.

Keywords: cognitive skills; desktop display; Gear VR; mental rotation test; mental cutting test; purdue spatial visualization test; spatial ability; virtual environments; virtual reality

1   Introduction

Spatial ability is an important skill in the modern world as engineering applications; even jobs require a good use of it. This ability allows the person to understand spatial relations between objects and space. As it is a cognitive ability,

it can be measured and improved by solving simple geometric problems. These geometric problems were recreated as tests and they have existed since the last century. These tests are called the Mental Rotation Test (MRT) [1], the Mental Cutting Test (MCT) [1, 2] and the Purdue Spatial Visualization Test (PSVT) [1, 3]. Each test type has a different type of mental rotation.

To this day, these tests mostly exist in paper-based formats. However, humanity is transitioning into a digital age, thus computers are spreading into the field of education: Virtual reality (VR) can positively affect the learning skills of students [4, 5]. Also, with the inception of Cognitive InfoCommunications (CogInfoCom) [6-8], it becomes easier to investigate the relationship of the human, information and communication technologies (ICT) and therefore, new cognitive capabilities can emerge [9].

Normally, the state of the art of improving spatial ability is vast. Sadly, most methods to improve spatial ability only exist in paper-based formats or in reality, but not in VR. There are, however, some exceptions which are digitalized or VR versions of these methods. These exceptions are mentioned in this paper.

The authors of [10] implemented the Paper Folding Test in VR. In [11] they assessed gender differences in mental rotation and the spatial ability of users in VR. Similarly, in [12] they assessed the spatial ability of both males and females in VR and in augmented reality (AR). Both studies concluded that males are better in mental rotation than females in VR and in AR. Also, the latter study suggests that AR could be a good tool for improving spatial ability. In a pilot study in [13], it has been concluded that VR is an effective spatial ability improving tool.

The three newest studies in this field are [14-16]. Similarly to [13], in [14] they concluded that VR is effective for improving spatial ability. In [15], they used the PSVT-rotation (PSVT-R) test and concluded that with VR goggles the users showed a significant improvement over the ones who used a desktop display. The study seen in [16] is not based on these geometric tests, but on navigating in VR. In the study, they concluded that actively navigating in VR can improve the spatial ability of older users.

The authors of this article developed a VR application [17] that contains the MRT, MCT and PSVT tests. This application gathers the data of spatial skills of the users in real-time. After gathering the data, the authors can analyze it. The results of the analysis are presented in this paper.

This paper is structured as the following: In the next section, the authors set up the research questions and hypotheses. Section 3 deals with the methodology used in testing. Section 4 presents the results. In section 5, the hypotheses are accepted or rejected, then the theses are formed. Lastly, conclusions are made.

2   Research Questions and Hypotheses

During the application development phase and before the testing, the authors set up 11 research questions (RQ) and hypotheses (H). The first five are about the tests on the desktop display, the next five are about the tests on the Gear VR, and the last one is about comparing the devices. The RQs are the following:

- RQ1: Which test mode is the easiest and the hardest when using a desktop display?

- RQ2: Is there any difference between the male and female performances using a desktop display?

- RQ3: Does the primary hand of the user influence the results of the tests when using a desktop display?

- RQ4: Does the age of the user affect the results of the tests when using a desktop display?

- RQ5: Does the major of the user affect the results of the tests when using a desktop display?

- RQ6: Which test mode is the easiest and the hardest with the Gear VR?

- RQ7: Is there any difference between the male and female performances on the tests with the Gear VR?

- RQ8: Does the primary hand of the user influence the results of the tests when using the Gear VR?

- RQ9: Does the age of the user affect the results of the tests on the Gear VR?

- RQ10: Does the major of the user affect the results of the tests when using the Gear VR?

- RQ11: With which device do the users achieve better results on the tests?

As the statistical hypothesis testing test the equality, and the alternative hypothesis is the nonequality, the authors formulate the following hypotheses:

- H1: In the case of desktop display, the average rates of correct answers are the same in case of all types of tests, opposite to, they depend on the test type. In case of different values of average rates, are they the same with both display devices? Do the statistical evaluations reflect the subjective opinions of students: "The MCT mode is the hardest and the PSVT mode is the easiest when using a desktop display".

- H2: The performances of males and females are equal, opposite to males perform better on the tests when using a desktop display.

- H3: The performances of left-handed and right-handed people are equal, opposite to left-handed people perform better on the tests when using a desktop display.

- H4: The performances of older people are equal to younger people, opposite to older people perform better on the tests when using a desktop display.

- H5: The performances of the students with different major are equal, opposite to, they differ when using a desktop display.

- H6: The average rates of correct answers are the same in case of all types of test opposite to they depend on the test type when using the Gear VR. Do the statistical evaluations reflect the subjective opinions of students: "The MCT mode is the hardest and the PSVT mode is the easiest when using the Gear VR".

- H7: The performances of males and females are equal, opposite to males perform better on the tests when using the Gear VR.

- H8: The performances of left-handed and right-handed people are equal, opposite to left-handed people perform better on the tests with the Gear VR.

- H9: The performances of older people are equal to younger people, opposite to older people perform better on the tests with the Gear VR.

- H10: The performances of the students with different major are equal, opposite to, they differ when using the Gear VR.

- H11: The average rates of correct answers are equal if the user uses desktop display and Gear VR, opposite to the users who test with the Gear VR achieve better results.

3   Methodology

An application for the mentioned tests was developed in the Unity game engine (version 2018.3.14f1) at the University of Pannonia during the first half of 2019. It contains the MRT, MCT and PSVT tests, as seen in Figure 1. Each test has ten rounds of questions relating to spatial ability. The application runs on Windows operating systems and on Android due to the Gear VR device, namely the Samsung Galaxy Gear VR SM-R322 with a Galaxy S6 Edge+ smartphone.

The tests were conducted with the application at the University of Pannonia and at the University of Debrecen during September 2019. At the former, 61 students tested the application with the Gear VR device and at the latter, 240 students

tested the application with a desktop display device, namely with an LG 20M37A (19,5") display device. Students who tested with the Gear VR consisted of Information Technology (IT) and non-IT students. Those who tested with the desktop display were either Architectural Engineering (AE) or Mechanical Engineering (ME) students.
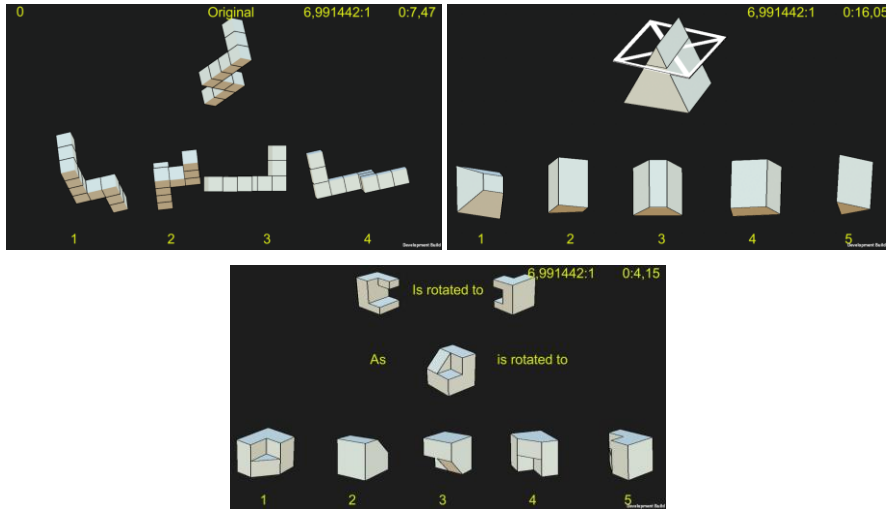

Figure 1
Screenshots of the MRT (left), MCT (right) and PSVT (center) test types

As one Gear VR device was available at the University of Pannonia, the VR testers came in sequential order through three weeks. The number of testers depended on the day, the least number of testers a day was 2 people and the most was 8 people. Testing at the University of Debrecen was different: The testers did the tests in a computer laboratory. There were twenty groups, each around 20 people. Each tester had to do all three types of spatial ability tests: The MRT, MCT and PSVT tests. After all three have been done, each tester had to do each type two more times. One tester lasted approximately 45 minutes.

The application logged the following information into a .csv file:

- The gender of the user, the age of the user, the primary hand of the user, the number of years spent at a university and what the user majored in.

- The test type, the test time, and the number of correct and incorrect answers.

- The application also logs technical information of each test, such as the virtual camera type, field of view, contrast ratio, rotation and whether shadows are turned on or off. This category is not focused on in this paper.

The ratios of the correct answers by the students were investigated. The authors checked the hypotheses of normal (Gauss) distribution by Kolmogorov-Smirnov tests. Then, the authors tested the equality of standard deviations (dispersions) and expectations. For the cases of dispersions, the authors applied F tests, and equality of expectations were checked by t-test or Welch-test [18]. The calculations were carried out by the statistical program package R.

4   Results

In this section, the results of the tests can be seen. The authors divided this section into three subsections: Subsection 4.1 deals with the desktop display results, subsection 4.2 with the Gear VR results, and subsection 4.3 is the comparison between the two devices.

4.1   Desktop Display Results

This subsection deals with desktop display results. Here, five different analyses can be found: The difficulty of the tests, the ratios of correct answers by gender, by primary hand, by age, and by majors.

4.1.1   Difficulty of the Tests

In Tables 1, 2 and 3, statistical data of all test types and even of their difficulties can be seen.

Table 1

Ratios of correct answers by test type with a desktop display

	Students	Min	Max	Average	Dispersion
MRT	240	0.4167	1	0.8041	0.1334
MCT	240	0.1333	0.9667	0.4389	0.154
PSVT	240	0.1333	0.9667	0.6168	0.1932

Table 2

Comparisons of standard deviations of the ratios of correct answers with a desktop display, also showing what is equal (eq)

	MRT			MCT			PSVT		
	Test value	Sign.	Eq.	Trial stat.	Sign.	Eq.	Test value	Sign.	Eq.
MRT	0	1	Yes	0.7503	0.0268	No	0.4764	0	No
MCT	0.7503	0.0268	No	0	1	Yes	0.6349	0.0005	No
PSVT	0.4764	0	No	0.6349	0.0005	No	0	1	Yes

Table 3

Comparisons of averages rates of correct answers with a desktop display, also showing what is equal (eq)

	MRT			MCT			PSVT		
	Test stat.	Sign.	Eq.	Test stat.	Sign.	Eq.	Test stat.	Sign.	Eq.
MRT	0	1	Yes	27.775	0	No	12.358	0	No
MCT	27.775	0	No	0	1	Yes	-11.155	0	No
PSVT	12.358	0	No	-11.155	0	No	0	1	Yes

As can be seen from the Tables 1, 2 and 3, every test type is distinguishable from the others. Also, the MCT test is the hardest type of test. The MRT type is the easiest and the PSVT test stands between MRT and MCT in terms of difficulty.

4.1.2 Comparison of Genders

The authors tested the type of distribution regarding genders. The hypothesis of normal distribution was accepted with p-value = 0.4846 in the case of males and with p-value = 0.9707 in the case of females. Table 4 shows the ratios of correct answers.

Table 4

Ratios of correct answers by gender with a desktop display

	Students	Min	Max	Average	Dispersion
Male	211	0.3083	0.9667	0.6769	0.1172
Female	29	0.4417	0.7833	0.5865	0.0966

The equality of dispersions is accepted with p-value = 0.2213, but the equality of average rates is rejected on the level of significance 0.00004. Thus, the ratio of correct answers is significantly better for males in the case of the desktop display.

4.1.3 Comparison of the Primary Hand of the User

Next, the effect of the primary hand of the users was investigated. The hypothesis of normal distribution was accepted for both primary hands. For right-handed users, it was accepted with p-value = 0.5343 and p-value = 0.9313 for the left-handed users. The ratios of correct answers of users in the case of the desktop display are found in Table 5.

Table 5

Ratios of correct answers by primary hand with a desktop display

	Students	Min	Max	Average	Dispersion
Right-handed	213	0.3083	0.9667	0.6691	0.1176
Left-handed	27	0.4417	0.8833	0.6414	0.1242

T Guzsvinecz *et al.*

Analyzing the Spatial Skills of University Students with a
Virtual Reality Application using a Desktop Display and the Gear VR

The equality of dispersions was accepted with p-value = 0.6567 and the equality of the expected values is also accepted with p-value = 0.2796. Therefore, there is no significant difference between right-handed and left-handed people when using a desktop display.

4.1.4 Comparison of age Groups

Afterward, an age-group analysis was made. The basic data can be seen in Table 6, and the analysis can be seen in Table 7.

Table 6

Statistical data of rates of correct answers by age groups with a desktop display

Age	Students in group	Group average	Group dispersion
17	1	0.666	0
18	33	0.623	0.124
19	89	0.667	0.116
20	75	0.680	0.111
21	29	0.673	0.125
22	6	0.573	0.087
23	2	0.750	0.087
24	2	0.612	0.205
25	1	0.808	0
27	1	0.758	0
32	1	0.866	0

Table 7

Comparing different age groups who used a desktop display

	Group 1	Avg. rate 1	Group 2	Avg. rate 2	p-value	Significant difference
<= 17 & > 17	1	0.6667	239	0.666	0.9277	No
<= 18 & > 18	34	0.6245	206	0.6728	0.0375	Yes
<= 19 & > 19	123	0.6556	117	0.6769	0.1656	No
<= 20 & > 20	198	0.6652	42	0.6696	0.8357	No
<= 21 & > 21	227	0.6662	13	0.6615	0.9047	No
<= 22 & > 22	233	0.6638	7	0.7369	0.1801	No
<= 23 & > 23	235	0.6646	5	0.7317	0.3885	No
<= 24 & > 24	237	0.6641	3	0.8111	0.0359	Yes
<= 25 & > 25	238	0.6647	2	0.8125	0.2179	No
<= 27 & > 27	239	0.6651	1	0.8667	0	Yes

As can be seen from Table 7, there are three significant differences in the age groups. There is a significant difference when comparing people who are less than or equal to 18 years of age and to those who are older. While there are similar

results in the "<=24 & >24" and "<= 27 & > 27" age groups, the number of people in those groups is small, thus the results may change if more people from those groups do the tests.

Comparison of Majors

The users who did the test on the display device can also be categorized into two majors: AE and ME. The ratios of their correct answers are found in Table 8. Normality analyses were executed. For AE, the p-value = 0.8103 and for ME, the p-value = 0.8763. The results of the analysis are accepted.

Table 8
Rates of correct answers by the major of the users using desktop display

	Students	Min	Max	Average	Dispersion
AE	62	0.4083	0.8500	0.6460	0.1127
ME	178	0.3083	0.9667	0.6729	0.1200

The equality of the standard deviations is accepted with p-value = 0.5774. The equality of average rates is also accepted with p-value = 0.1133. Therefore, as can be seen, there is no significant difference between the results of AE and ME.

4.2   Gear VR Results

This subsection presents the results of the users who used the Gear VR. Similarly, to subsection 4.1, the same five analyses can be found, but the data are from the users who used the Gear VR. The subsubsections contain information on the difficulty of the tests, correct answers by gender, primary hand, age, and majors.

4.2.1   Difficulty of the Tests

Firstly, the difficulty of the test types was examined. Similarly, to before the authors grouped the difficulties according to the test types.

Statistical data of the test type difficulties can be seen in Tables 9, 10 and 11. Table 9 shows the rates of correct answers by test type, Table 10 shows the standard deviation of the rates of correct answers and Table 11 compares the rates of correct answers.

Table 9
Rates of correct answers by test type with the Gear VR

	People	Min	Max	Average	Dispersion
MRT	61	0.4833	0.9833	0.8003	0.1248
MCT	61	0	0.8000	0.4071	0.1722
PSVT	61	0.0667	1	0.72295	0.1886

Table 10

Standard deviations of rates of correct answers with the Gear VR, showing what is equal (eq)

	MRT			MCT			PSVT		
	Test stat.	Sign.	Eq.	Test stat.	Sign.	Eq.	Test stat.	Sign.	Eq.
MRT	1	1	Yes	0.5247	0.0136	No	0.4376	0.0016	No
MCT	0.5247	0.0136	No	1	1	Yes	0.8340	0.4842	Yes
PSVT	0.4376	0.0016	No	0.8340	0.4842	Yes	1	1	Yes

Table 11

Comparison of average rates of correct answers with the Gear VR, also showing what is equal (eq)

	MRT			MCT			PSVT		
	Test stat.	Sign.	Eq.	Test stat.	Sign.	Eq.	Test stat.	Sign.	Eq.
MRT	0	1	Yes	14.437	0	No	2.6704	0.0087	No
MCT	14.437	0	No	0	1	Yes	-9.657	0	No
PSVT	2.6704	0.0087	No	-9.657	0	No	0	1	Yes

Similarly, to the desktop display, every test mode is distinguishable. However, with the Gear VR, comparing only the dispersions, in some cases there are no, but in other cases, there are significant differences between them. By investigating the average rates of correct answers, the authors concluded that there are significant differences between all test types. The difficulty is the same as the desktop display results: MCT mode is the most difficult while MRT mode is the easiest.

4.2.2 Comparison of Genders

Out of the 61 users who performed the tests, 44 were male users and 17 were female users. Therefore, the comparison was done regarding the gender of the user. The number of correct answers is found in Table 12. A normality analysis was done with p-value = 0.5377 for males and with p-value = 0.6657 for females.

Table 12

Rates of correct answers by gender using the Gear VR

	Students	Min	Max	Average	Dispersion
Male	44	0.3667	0.9000	0.6790	0.1247
Female	17	0.4417	0.8000	0.6922	0.1092

The equality of the dispersions is accepted, (p-value = 0.5757), and the equality of average rates is also accepted at a high-level of significance (p-value = 0.6875). Therefore, there is no significant difference between the two genders.

However, this nonexistence of the significant difference is a different result than the result with the desktop display: In the previous subsection, the result was that

the male users performed significantly better than the female users when using a desktop display. This means that by using the Gear VR, the females performed better than their desktop display counterparts. More information on this different result is available in the next subsection.

4.2.3   Comparison of the Primary Hand of the User

Next, a comparison was made regarding the primary hand of the users. A normality analysis was done, the hypothesis of Gauss distribution is accepted at high levels of significance (p-value = 0.3623 for the right-handed and p-value = 0.9937 for the left-handed users).

Table 13

Rates of correct answers by primary hand with the Gear VR

	Students	Min	Max	Average rate	Dispersion
Right-handed	52	0.3667	0.8583	0.6922	0.1193
Left-handed	9	0.6000	0.9000	0.7611	0.0938

The equality of the standard deviations is accepted (p-value = 0.4826), but the equality of average rates is rejected on the level of significance 0.05 (p-value = 0.02201.

Thus, by using the Gear VR, the left-handed users performed significantly better than their right-handed counterparts. This result is different from the last subsection, as when using a desktop display, there was no significant difference between the performance of left-handed and right-handed users.

This means that the Gear VR not only improved on the performance of the female users but on the performance of left-handed users as well. More information on these different results is available in the next subsection.

4.2.4   Comparison of Age Groups

After the data regarding the primary hand of the user was evaluated, the next data to analyze is the age groups. For information regarding the different age groups, see Tables 14 and 15.

In Table 14, the statistical data of correct answers using the Gear VR is presented and grouped by age groups. The number of people, the group average and the group dispersion can be seen.

In Table 15, the different age groups who took the tests with the Gear VR are compared and examined: The "Less than and equal to" certain age groups were compared to "greater than" certain age groups. The results are the following:

Table 14

Statistical data of correct answers by age groups with the Gear VR

Age	Number of students in the group	Average rate	Dispersion
19	3	0.7306	0.1008
20	4	0.6	0.1763
21	9	0.6296	0.1176
22	10	0.6667	0.1153
23	11	0.672	0.1492
24	8	0.6938	0.1006
25	4	0.725	0.1369
26	3	0.6944	0.0376
27	3	0.75	0.1228
28	2	0.7542	0.0412
30	1	0.8083	0
31	1	0.6583	0
32	1	0.8083	0
34	1	0.7583	0

Table 15

Comparing different age groups who used the Gear VR

Age	Group 1	Avg. rate 1	Group 2	Avg. rate 2	p-value	Significant difference
<= 19 & > 19	3	0.7306	58	0.6802	0.4812	No
<= 20 & > 20	7	0.656	54	0.6861	0.633	No
<= 21 & > 21	16	0.6411	45	0.6974	0.1396	No
<= 22 & > 22	26	0.651	35	0.7062	0.0792	No
<= 23 & > 23	37	0.6572	24	0.7219	0.0266	Yes
<= 24 & > 24	45	0.6637	16	0.6637	0.0163	Yes
<= 25 & > 25	49	0.6687	12	0.6687	0.0155	Yes
<= 26 & > 26	52	0.6702	9	0.6702	0.014	Yes
<= 27 & > 27	55	0.6745	6	0.6745	0.0157	Yes
<= 28 & > 28	57	0.6773	4	0.6773	0.0996	No
<= 30 & > 30	58	0.6796	3	0.6796	0.2915	No
<= 31 & > 31	59	0.6792	2	0.6792	0.0753	No
<= 32 & > 32	60	0.6814	1	0.6814	0	Yes

As can be seen from Table 15 above, the users who are 23-year-old's or are older, tested with the Gear VR have achieved significantly better results than the ones who are younger.

4.2.5   Comparison of Majors

Since the users who tested the application majored in different areas than the ones in Debrecen, the authors set up a new category. When comparing the "majored in" areas, the authors set up two categories: IT students and non-IT students. For information, see Table 16 which presents the rates of correct answers.

Table 16

Rates of correct answers by the major using Gear VR

	Students	Min	Max	Average	Dispersion
IT	21	0.4417	0.8667	0.6845	0.1103
Non-IT	40	0.3667	0.9000	0.6817	0.1259

The hypotheses of Gauss distribution were accepted (p-value = 0.9854 for IT students and p-value = 0.2599 for non-IT students). The equality of the dispersions is accepted (p-value = 0.5338), and so is the equality of the average rates (p-value = 0.9275). Therefore, there is no significant difference concerning the spatial skills measured by our tests between IT students and non-IT students.

4.3   Comparison of Gear VR and Desktop Display Device Results

In this subsection, the authors compare the results of the desktop display and the Gear VR. A comparison is done with the same statistical data as seen in the subsections above. This means the difficulty of the tests, correct answers by gender, by primary hand, and by age. The authors wanted to compare the results by majors, but due to different majors at the universities, it could not be done.

4.3.1   Difficulty of the Tests

The difficulty of the tests was compared in each case. The data can be seen in the following table. D means desktop display, VR means the Gear VR. 1 is the MRT test, 2 is the MCT test, and 3 is the PSVT test. Table 17 presents the rates of correct answers by test type using different display devices.

Table 17

Rates of correct answers by test type with each display device

Device	Students	Min	Max	Average rates	Dispersion
D1	240	0.4167	1	0.8041	0.1334
VR1	61	0.4833	0.9833	0.8003	0.1248
D2	240	0.1333	0.9667	0.4389	0.154
VR2	61	0	0.8	0.4071	0.1723
D3	240	0.1333	0.9667	0.6168	0.1932
VR3	61	0.0667	1	0.723	0.1886

T Guzsvinecz *et al.*

Analyzing the Spatial Skills of University Students with a
Virtual Reality Application using a Desktop Display and the Gear VR

The equality of the standard deviations is accepted in cases D1-VR1 (p-value = 0.5476), D2-VR2 (p-value = 0.2454), D3-VR3 (p-value = 0.8456). The equality of average rates is accepted in the first two cases only: D1-VR1 (p-value = 0.8336) and D2-VR2 (p-value = 0.1924). In case of D3-VR3, the hypothesis of equality is rejected (p-value = 0.0001766), thus the Gear VR users performed significantly better performing the PSVT tests when compared with the desktop display devices.

4.3.2 Ratio of Correct Answers

Statistical data of the ratio of correct answers with each display device can be seen in Table 18. This ratio is illustrated in Figure 2 with cumulative distribution functions.

Table 18

Rates of correct answers with each display device

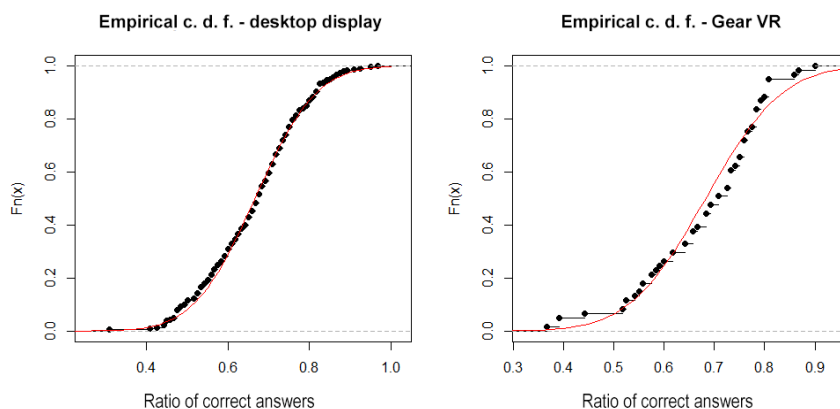|	Students	Min	Max	Average rate	Standard deviation
Desktop display	240	0.3083	0.9667	0.6660	0.1185
Gear VR	61	0.3667	0.9000	0.6827	0.1199

Figure 2

Ratio of correct answers for desktop displays (left) and the Gear VR (right)

Normality analyses were done for both devices. The p-value is 0.6335 for desktop displays and the p-value is 0.2548 for the Gear VR, hence the hypotheses of Gauss distributions were accepted in both cases.

The hypotheses of the equality of the dispersions and the expectations are accepted (p-value is 0.8786, and p-value = 0.3332, respectively) thus there is no significant difference of the correct answers between the two devices.

4.3.3   Comparison of Genders

Table 19

Rates of correct answers by gender using the different display device

	Students	Min	Max	Average	Dispersion
Desktop display, Male	211	0.3083	0.9667	0.6769	0.1172
Gear VR, Male	44	0.3667	0.9000	0.6790	0.1247
Desktop display, Female	29	0.4417	0.7833	0.5865	0.0965
Gear VR, Female	17	0.4417	0.8000	0.6922	0.1092

When comparing the male desktop display testers to the male Gear VR testers and with the data from Table 19 (illustrated in Figure 3), the equality of the standard deviations (p-value = 0.5609) and the averages (p-value = 0.9193) are both accepted. Moreover, the distribution of the two groups is tested, the equality of the distributions is accepted (p-value is 0.7931). Therefore, there is no significant difference between the two male groups. However, the result is different between the two female groups. When looking at the standard deviations, the equality is accepted with p-value = 0.5513, but when looking at the average rates, the equality is rejected with (p-value = 0.0024). This means that there is a significant improvement in the Gear VR female group. When the equality of the two distributions is tested, it is refused with p-value = 0.02125.

4.3.4   Comparison of the Primary Hand of the Users

Next step was to analyze based on the primary hands of the users.

Table 20

Rates of correct answers by primary hand with each display device

	Students	Min	Max	Average rate	Standard deviation
Desktop display, Right-handed	213	0.3083	0.9667	0.6691	0.1177
Gear VR, Right-handed	52	0.3667	0.8583	0.6691	0.1193
Desktop display, Left-handed	27	0.4417	0.8833	0.6414	0.1242
Gear VR, Left-handed	9	0.6000	0.9000	0.7611	0.0938

A comparison was done between the two right-handed groups, see Table 20 and Figure 4 for illustration. The equality of the standard deviations and also of the average rates are accepted on very high levels (p-value = 0.8633 and p-value = 0.9991, respectively) and the equality of the distributions is also accepted (p-value = 0.7086). Therefore, there is no significant difference between the two right-handed groups. The result is different between the two left-handed groups: The equality of the standard deviations is accepted (p-value = 0.4164), but the equality of average rates is rejected (p-value = 0.006949). This means that there is a significant improvement with the use of Gear VR in the case of the left-handed group.

4.3.5 Comparison of Age Groups

When doing a comparison between the two devices, a significant difference was found in age comparison. To achieve this result, four age groups were made: Users who tested with a desktop display device and are less than or equal to 18 years of age (DU18), or over 18 (DO18). The remaining two groups are users who tested with the Gear VR device and are less than or equal to 23 years of age (VRU23) or over 23 (VRO23).

The age groups were divided at these ages, as there were differences that could be seen. These data can be found in Table 21.

Table 21

Statistical data of correct answers by age groups with each display device

	Students	Min	Max	Average rates	Dispersion
DU18	34	0.3083	0.8583	0.6245	0.1224
VRU23	37	0.3667	0.9	0.6572	0.1297
DO18	206	0.4083	0.9667	0.6728	0.1167
VRO23	24	0.5417	0.8667	0.7219	0.0921

The equality of standard deviations is accepted (p-value = 0.739) in category DU18, VRU23, and is also accepted in category DO18 and VRO23 (p-value = 0.1794). The equality of expected rate values is accepted (p-value = 0.2784) in category DU18, VRU23. However, there is a significant difference between DO18 and VRO23 categories (p-value 0.02259). Therefore, the users who are over 23 and tested with the Gear VR are significantly better than the users who tested with the display device and are over 18.

5 Discussion

Since spatial ability is an important skill, tests were conducted. Data are collected by using the application and the collected data are analyzed. This produced the statistical results, as can be seen in multiple tables and figures in the previous section. Therefore, from the results, the authors accept hypotheses H3, H5, H7, H10, and H11. The answers for hypotheses H1 and H6 are mixed. The first half of them is rejected, the second half is accepted. The remaining hypotheses H2, H4, H8, H9 are fully rejected.

5.1 Accepted Hypotheses

Since left-handed users use the right side of their brain more often, the authors suspected that left-handed people perform better on the tests (alternative

hypotheses of H3 and H8). However, according to Tables 5, 13 and 20, H3 is accepted, forming T3: There is no significant difference between the performances of left-handed and right-handed people who used a desktop display.

Concerning H5 and H10, the statistics can be seen in Tables 8 and 16. Both H5 and H10 can be accepted. H5 forms T5: There is no significant difference in the performance of architectural engineering and mechanical engineering students when using a desktop display. Similarly, to H5, H10 forms T10: There is no significant difference in performance between IT and non-IT students when using the Gear VR.

Since the tests were conducted at two different universities, the majors of the students differ. However, the authors believe after forming T5 and T10 that the majors of the users do not influence their performance.

Concerning H7, by using the Gear VR, there is no significant difference between the performances of males and females. This is a different result in comparison to the result of the desktop display. This forms the following T7: Female users perform better numerically on the spatial ability tests than males by 1.94% on average with the Gear VR.

Taking RQ11 and H11 into account when formulating thesis T11, it can be seen from Table 18 that when using the Gear VR, the users can produce better results on the tests. Thus, H11 is accepted. This fact yields thesis T11: There is no significant difference in the ratio of correct answers when comparing desktop displays with the Gear VR, but with the latter, the users produced numerically a better average of correct results by 2.5%.

5.2   Mixed Cases

As mentioned before, hypotheses H1 and H6 present mixed cases. What can easily be seen from Tables 1, 2, 3, 9, 10, 11 and 17, that the users found the MCT test mode the hardest and the MRT test mode the easiest. Therefore, the first half of both H1 and H6 is rejected, and the second half is accepted. H1 forms thesis T1: The users who used a desktop display / Gear VR found the MCT test mode the hardest, and the MRT mode the easiest.

Comparing the devices, there is a new piece of information available in Table 17: Users of the Gear VR performed significantly better on the PSVT tests than their counterparts who used a desktop display. Therefore, T6 is the following: While there are no significant changes in the ratio of correct answers in the MRT, MCT test mode when comparing the desktop display to the Gear VR, the users who tested with the latter performed significantly better (17.217%) in the PSVT test mode.

5.3   Rejected Hypotheses

Investigating H2, Tables 4, 12 and 19 are taken into account. Table 4 shows the desktop display results, Table 12 shows the Gear VR results and Table 19 compares both of them. According to the mentioned tables, T2 is formed: When using a desktop display, there is a significant difference between males and females, meaning males perform better on the tests than females by 15.41% on average. This fact is similar to older studies not featuring VR, such as [19] where it is proven that males have better spatial skills than females. When using the Gear VR, however, this significant difference disappears, meaning that the females perform better on the tests than males by 1.94% on average. This means that the Gear improves especially the women's achievements, which is an interesting result.

H4 is rejected, as can be seen from Table 7. This formulates T4: The users who used a desktop display and are over 18 years of age performed better by 7.73% on average than their younger counterparts.

H8 is also rejected. With H8, T8 can be formed: There is a significant difference in the performance of the left-handed and right-handed people who used the Gear VR. The difference is quite large, it is about 7%. This means that the performance of the left-handed people was increased significantly by the Gear VR, as Section 4.3.4 shows, the increment is about 13%. This result is quite different from older studies not featuring VR, such as [20] where they concluded that right-handed people outperformed left-handed people. It seems like using a desktop display makes their performance equal and when using the Gear VR, left-handed people outperform right-handed ones.

Similarly, to H4, H9 is rejected as well. Data for this can be seen in Table 15. T9 is formed: The users who used the Gear VR and are over 23 years of age performed significantly better than their younger counterparts. (The difference is 9.4% numerically.) Comparing displays, Gear VR users who are over 23 years of age performed significantly better than the users who are over 18 years of age and used a desktop display (the difference is 7.29% on average).

Conclusions

A well-developed spatial ability is an important cognitive skill in the modern age. In the last century, only paper-based methods were available, but with the coming of the digital world, the number of methods is increasing. With virtual environments in virtual worlds, new methods can be created, or existing methods can be improved upon.

The CogInfoCom supported education has several advantages to improve the effectiveness of learning. For example, through the benefits and opportunities of Virtual Labs [21], gamification or team-based collaborative education [22, 23,

24], Human-Computer Interfaces [25, 26, 27], Virtual Reality-based learning spaces [28] and other educational environments [29, 30].

The authors of this paper created such a VR application that presents three old methods – MRT, MCT and PSVT – in a new context. This application allows to get data in real-time from the spatial ability tests and is also available on two platforms: PC and Android.

This application was tested by 61 students on the Android platform with the Gear VR and by 240 students on the PC platform with an LG desktop display. Data were gathered from the users and were analyzed. The data consisted of the age of the user, gender of the user, primary hand of the user, the number of years spent at the university and what the student majored in.

Eleven research questions and hypotheses were made in the beginning. Five theses were formed from five accepted hypotheses, four were formed from rejected hypotheses and two from mixed hypotheses.

With these eleven theses, it can be concluded that the Gear VR offers a significant improvement to the spatial performance of female users, left-handed users, and older users. With the use of the Gear VR, the PSVT test mode also becomes significantly easier to the users.

According to the results, the use of a VR headset positively influences the spatial skills of the users. This is good, as most education for engineers at universities contain subjects such as technical representation or descriptive geometry and a well-developed spatial ability is necessary for successful studies.

In conclusion, it can be safely stated that the use of virtual environments and virtual reality can help with developing and improving the spatial skills of students. Also, these results strengthen the fact that virtual reality has a future in education.

Acknowledgement

References

[1]	Ault, H. K., John, S.: Assessing and Enhancing Visualization Skills of Engineering Students in Africa: A Comprehensive Study. Engineering Design Graphics Journal, Vol. 74, No. 2, 2010, pp. 12-20

[2]	Bosnyák, Á., Nagy-Kondor, R.: The spatial ability and spatial geometrical knowledge of university students majored in mathematics, Acta Didactica Universitatis Comenianae, Vol. 8, 2008, pp. 1-25

[3]   Branoff, T., Connolly, P.: The Addition of Coordinate Axes to the Purdue
        Spatial Visualization Test – Visualization of Rotations: A Study at Two
        Universities, Proceedings of the American Society for Engineering
        Education Annual Conference, 1999, pp. 1-9

[4]   Kovari, A.: CogInfoCom Supported Education: A review of CogInfoCom
        based conference papers, Proceedings of the 9th IEEE International
        Conference on Cognitive Infocommunications (CogInfoCom), 2018. pp.
        233-236

[5]   I. Horváth: Innovative engineering education in the cooperative VR
        environment, 7th IEEE Conference on Cognitive Infocommunications,
        Wrocław, Poland, 2016, pp. 359-364

[6]   Baranyi, P., Csapó, Á., Balázs, Á., Várlaki, P.: An Overview of Research
        Trends in CogInfoCom, 18th International Conference on Intelligent
        Engineering Systems, 2014, pp. 181-186

[7]   Baranyi, P., Csapo, A., Sallai, Gy.: Cognitive Infocommunications
        (CogInfoCom), Springer International Publishing Switzerland (978-3-139-
        19608-4), 2015, p. 219

[8]   A. Csapo et al.: „VR as a Medium of Communication: from Memory
        Palaces to Comprehensive Memory Management", 9th IEEE International
        Conference on Cognitive Infocommunications, Budapest, 2018

[9]   Heldal, I., and Helgesen C.: The Digital HealthLab: Supporting
        Interdisciplinary Projects in Engineering and in Health Education, Journal
        of Applied Technical and Educational Sciences, Vol. 8, No. 4, 2018, pp. 4-
        21

[10]   Hercegfi, K., Komlodi, A., Szabo, B., Koles, M., Logo, E., Hamornik, B.
        P., Rozsa, Gy.: Experiences of virtual desktop collaboration experiments,
        6th IEEE International Conference on Cognitive Infocommunications, 2015,
        pp. 375-379, DOI: 10.1109/CogInfoCom.2015.7390622

[11]   Parsons, T. et al.: Sex differences in mental rotation and spatial rotation in a
        virtual environment. Neuropsychologia, Vol. 42, No. 4, 2004, pp. 555-562,
        DOI: 10.1016/j.neuropsychologia.2003.08.014

[12]   Dünser, A., Steinbügl, K., Kaufmann, H., & Glück, J.: 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, 2006, pp. 125-132

[13]   Torner, J., Alpiste, F., Brigos, M.: Virtual Reality application to improve
        spatial ability of engineering students. WSCG 2016 - 24th Conference on
        Computer Graphics, Visualization and Computer Vision 2016

[14]   Molina-Carmona, R., Pertegal-Felices, M. L., Jimeno-Morenilla, A., &
        Mora-Mora, H.: 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, 2018, pp. 171-185

[15] Molina-Carmona, R., Pertegal-Felices, M. L., Jimeno-Morenilla, A., Mora-Mora, H. (2018): Virtual Reality Learning Activities for Multimedia Students to Enhance Spatial Ability. Sustainability, Vol. 10, No. 4, 2018, DOI: 10.3390/su10041074

[16] Meade, M. E., Meade, J. G., Sauzeon, H., & Fernandes, M. A.: Active Navigation in Virtual Environments Benefits Spatial Memory in Older Adults. Brain sciences, Vol. 9, No. 3, 2019, DOI: 10.3390/brainsci9030047

[17] Guzsvinecz, T., Szeles, M., Perge, E., Sik-Lanyi, C.: Preparing spatial ability tests in a virtual reality application, 10th IEEE International Conference on Cognitive InfoCommunications, 23-25 October 2019, Naples, Italy, pp. 363-367

[18] Prékopa, A.: Probability theory of engineering applications (Valószínűségelmélet műszaki alkalmazásokkal). Műszaki Kiadó, 1974, Budapest

[19] Burnett, S. A., Lane, D. M., Dratt, L. M.: Spatial ability and handedness. Intelligence, Vol. 6, No. 1, 1982, 57-68

[20] McKeever, W. F.: The influences of handedness, sex, familial sinistrality and androgyny on language laterality, verbal ability, and spatial ability. Cortex, Vol. 22, No. 4, 1986, 521-537

[21] Budai, T., Kuczmann, M.: Towards a Modern, Integrated Virtual Laboratory System, Acta Polytechnica Hungarica, Vol. 15, No. 3, 2018, pp. 191-204

[22] Rigóczki, Cs., Damsa, A., Györgyi-Ambró, K.: Gamification on the edge of educational sciences and pedagogical methodologies, Journal of Applied Technical and Educational Sciences, Vol. 7, No. 4, 2017, pp. 79-88

[23] Fürstner, I., Gogolák, L., Sarcevic, P.: Development of telepresence technology during the teaching process at Subotica Tech, Journal of Applied Technical and Educational Sciences, Vol. 8, No. 4, 2018, pp. 44-53

[24] Pinter, R., Cisar, S. M.: Measuring Team Member Performance in Project Based Learning, Journal of Applied Technical and Educational Sciences, Vol. 8, No. 4, 2018, pp. 22-34

[25] Sziladi, et al. The analysis of hand gesture based cursor position control during solve an IT related task, 8th IEEE International Conference on Cognitive Infocommunications, 2017, pp. 413-418

[26] Jozsef Katona et al.: Electroencephalogram-Based Brain-Computer Interface for Internet of Robotic Things, in Cognitive Infocommunications, Theory and Applications, 2018, pp. 253-275

[27]   Costescu Cristina, et al.: Assessing Visual Attention in Children Using GP3
       Eye Tracker, Proceedings of the 10th IEEE International Conference on
       Cognitive Infocommunications, 2019, pp. 343-348

[28]   Bognar, L., et al.: Improved learning environment for calculus courses,
       Journal of Applied Technical and Educational Sciences, Vol. 8, No. 4,
       2018, pp. 35-43

[29]   Ujbanyi, T., et al.: ICT Based Interactive and Smart Technologies in
       Education-Teaching Difficulties. International Journal of Management and
       Applied Science, 2017, Vol. 3, No. 10: pp. 72-77

[30]   Gogh, E., et al.: Metacognition and Lifelong Learning: A survey of
       secondary school students, 9th IEEE International Conference on Cognitive
       Infocommunications, 2018, pp. 271-276