Comparing the Results Received on Desktop Display and Paper Versions of Four Spatial Ability Test Types

Tibor Guzsvinecz[1], Eva Orban-Mihalyko[2], Cecilia Sik-Lanyi[3], Erika Perge[4]

[1]Department of Information Technology and its Applications, Faculty of Information Technology, University of Pannonia,
Gasparich M. utca 18/A, 8900 Zalaegerszeg, Hungary

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

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

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

[1]guzsvinecz.tibor@zek.uni-pannon.hu, [2]orban.eva@mik.uni-pannon.hu, [3]lanyi.cecilia@mik.uni-pannon.hu, [4]perge@eng.unideb.hu

Abstract: The spatial skills of freshman engineering students in higher education are investigated in this paper. For the analysis, the Mental Rotation Test, Mental Cutting Test, Purdue Spatial Visualization Test, and Heinrich Spatial Visualization Test were used. These tests were done on paper and in a non-immersive virtual environment using a desktop display by 201 and 205 students, respectively. The evaluation was done in the statistical program package R. According to the results, a better average of the ratio of correct answers is received on paper by 10.41%. To understand this difference, the rates of correct answers were also grouped and evaluated by test type, gender, dominant hand, age, and current studies of the participants. According to the results by test type, significant increases between the two versions occur on the Purdue Spatial Visualization Test, and the Synthesis as well as Decomposition subtypes of the Heinrich Spatial Visualization Test by 10.99%, 12.35%, and 22.07%, respectively. Besides this fact, each group of participants performed significantly better on paper than in non-immersive virtual environments.

Keywords: heinrich spatial visualization test; mental cutting test; mental rotation test; purdue spatial visualization test; spatial ability

1 Introduction

According to the Theory of Multiple Intelligences, spatial intelligence is one of nine that humans have [1]. Spatial intelligence can be defined as “the ability to generate, retain, retrieve, and transform well-structured visual images” [2]. As such, it allows the person to think in three dimensions. Spatial ability is also related to spatial intelligence, as it could be defined as contextual spatial intelligence that is real, and occurs in everyday life [3]. It is also a cognitive skill [4], and it is originally made of the concept of three definitions: mental rotation, spatial perception, and visualization [1]. According to Zacks [5], the former activates areas of the brain that involves motor stimulation. In the study of Jeannerod and Jacob [6], it is shown that the middle one involves the human visual system and the parietal lobe in the brain. Lastly, according to Motes et al. [7], the brain activities of people who used the latter were greater in the lateral occipital complex, the right superior parietal, dorsolateral prefrontal cortex as well as in the right ventrolateral prefrontal cortex than of objects visualizers. Later, the definition of spatial ability was expanded to the concept of five definitions by Maier [8]: spatial perception, visualization, mental rotation, spatial relations as well as rotations.

According to Ghiselli, spatial skills are essential in the fields of mathematics, architecture, and engineering [9]. This is due to the fact that many tasks in these fields require design and creation. It was also suggested that the curriculum of engineering studies should be extended with spatial ability improving subjects [10]. Since then, this suggestion was accepted, and universities began to train the spatial skills of their engineering students with subjects such as technical representation and descriptive geometry. In the end, according to the conclusions of Peters et al. [11], the spatial skills of engineering students were greater than that of non-engineering students. Also, the spatial skills of males are usually in favor when compared to females [12]. Differences between them can be seen as early as the age of seven and eight [13]. However, according to Brownlow and Miderski, the difference between the spatial skills of males and females is smaller and their results are converging to each other in the modern age [14].

Since spatial ability is a cognitive skill and not a biological susceptibility, it can be improved [15]. Originally, it begins to develop during childhood over three stages [16]. The development of spatial perception starts at the age of nine. It is important to note that age affects spatial ability: it improves in childhood, but declines in adulthood [17, 18]. However, education can improve it starting from the age of nine [19]. Besides learning environments [20], recreational activities can also train spatial skills [21]. Geometric problems on paper were developed through the years to measure and train the spatial ability of people. Even though their number is large, four are focused on in this paper. These four are the Mental Rotation Test (MRT) [22], Mental Cutting Test (MCT) [23], Purdue Spatial

Visualization Test (PSVT) [24], and the Heinrich Spatial Visualization Test (HSVT) [25].

As can be seen, these tests can assess cognitive skills. If we want to assess these in the digital world, the Cognitive InfoCommunications (CogInfoCom) field of research provides the perfect toolkit for this study. CogInfoCom is an interdisciplinary field that examines the relationship between human, information, and communication technologies [26]. Thus, it gives an opportunity to investigate several human factors using modern cognitive IT methods. Among others, human-computer interaction as well as human vision are investigated in the field of CogInfoCom [27-29].

Several authors concluded that the use of immersive virtual reality (VR) technologies can either improve the spatial skills of users or they simply can receive better results on the tests [30-32]. However, what about non-immersive ones such as a desktop display? Due to this, the following research question is asked: is there a difference between the results on paper and using a desktop display? Therefore, this study uses cognitive information technology methods, using a VE developed in Unity to examine the differences between traditional paper-based and non-immersive technology-based spatial ability assessments.

As was mentioned earlier in the paper, subjects that improve spatial skills are now included in the curriculum of engineering studies. Such is the case with engineering education at the University of Debrecen where one of the main focuses is on technical drawings. One of these subjects called descriptive geometry and it can be found in the first semester of certain engineering students. However, how developed are the spatial skills of freshman engineering students when they enter higher education?

To answer this research question, the spatial skills of students were measured on the aforementioned four test types and are evaluated as well as presented in this paper. The structure of the paper is as follows: the materials and methods are presented in Section 2, while the results are shown and discussed in Section 3. Conclusions are made in Section 4.

2 Materials and Methods

The four test types were recreated on paper and in a virtual environment (VE) developed in 2019 using Unity, compatible with both desktop displays and the Gear VR headset. Initially, the VE included the MRT, MCT, and PSVT tests, and was expanded with both HSVT sub-types in 2021. This study, however, used only the paper and desktop display versions. Details of the test types are provided in Sections 2.1-2.4, with data collection and analysis described in Sections 2.5-2.6.

2.1 The Mental Rotation Test

This test type was developed by Shepard and Metzler [22]. During the MRT test, an object is shown at the top of the paper/screen. This object has to be rotated mentally. Four possible answers exist, shown at the bottom of the paper/screen. These possible answers are already rotated. However, not one, but two of them are identical objects to the one above them (due to the rules of the MRT test). In our version, ten MRT questions were asked of the students. Figure 1 shows the paper and virtual versions of our implementation of the MRT test. In the case of the desktop display version, the objects could be selected by pressing the left mouse button when the cursor is on them, or by pressing 1-4 on the keyboard.

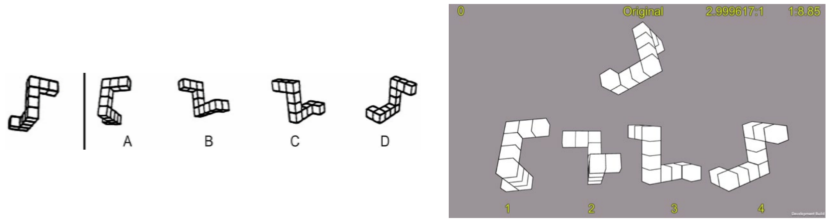

Figure 1

The MRT test in two implementations: Our paper (left) and non-immersive versions of it (right)

2.2 The Mental Cutting Test

The MCT test was developed by the College Entrance Examination Board [23]. This test type also has to be solved mentally. The test looks like the following: as before, an object is shown at the top of the paper/screen. However, in the case of this test, the cross-section of a solid and an object has to be recognized. Thus, five possible cross-sections are shown. These are the possible answers to this test, and only one of them is correct. In our version, ten MCT questions were asked. Our implementations of these test type can be seen in Figure 2. Similarly to the MCT test, object selection in the same fashion with the keyboard between 1-5.

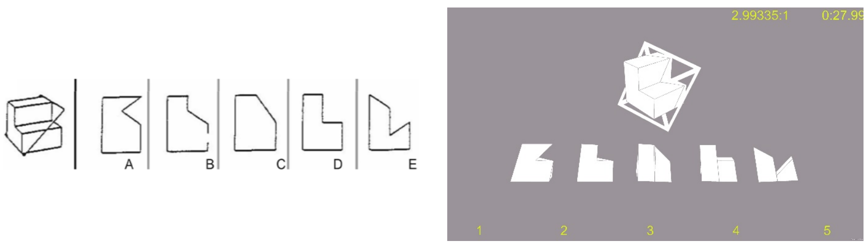

Figure 2

The MCT test in two implementations: Our paper (left) and non-immersive versions of it (right)

2.3 The Purdue Spatial Visualization Test

The PSVT was proposed by Guay [24]. As with the previous tests, this one also has to be solved mentally. In this case, however, there are three objects in the top of the paper/screen. The first two are example objects and these ones show a form of rotation. This has to be understood. Then, the third object has to be rotated in the same direction as before. There are five possible answers to this test, and each is identical to the third object. The correct rotation has to be selected. Ten PSVT questions were asked in our implementation, one of which is shown in Figure 3. As before, object selection was similar in the case of the desktop display version.

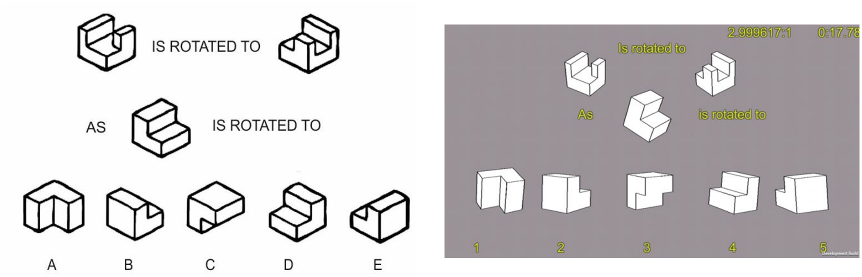

Figure 3
The PSVT test in two implementations: Our paper (left) and non-immersive versions of it (right)

2.4 The Heinrich Spatial Visualization Test

Lastly, the HSVT was developed by Heinrich [25]. This test type was created to focus on two specific skills. Therefore, it has two subtypes: synthesis and decomposition. When the synthesis version is done, the objects on the test have to be fitted together mentally. The participants have to choose one of the five possible answers, and this object has to fit between X and the end piece. If all objects in the center can fit together, a correct answer is chosen. In the case of the decomposition version, there are five possible answers and one correct one that have to be chosen. This chosen object is needed to fit mentally between X and Y. Thus, the object that is on the left of the arrow could be constructed by adding X, the chosen one, and Y together mentally. Our implemented versions of both HSVT subtests can be seen in Figure 4. The chosen object is symbolized by a question mark in the case of each implementation. As before, object selection was similar in the case of the desktop display version.

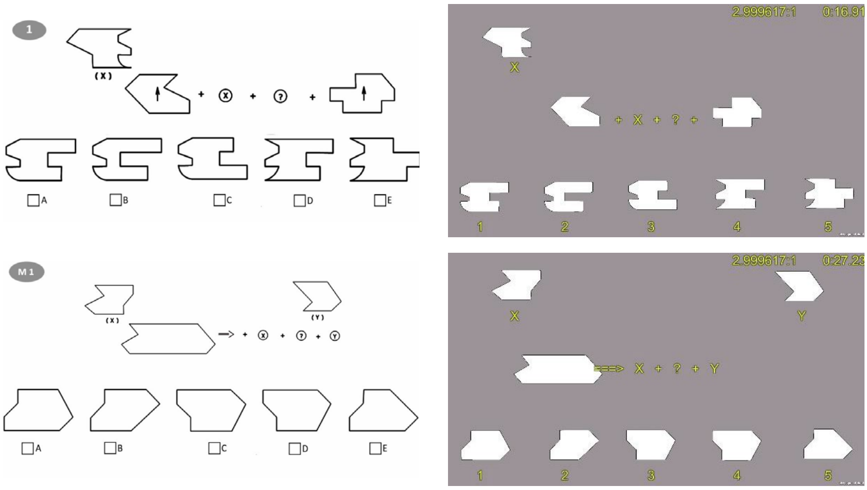

Figure 4
The two subtests of HSVT test in two implementations: Our paper (upper left) and non-immersive versions of the HSVT – Synthesis (upper right). Our paper (lower left) and non-immersive versions of the HSVT – Decomposition (lower right).

2.5 Data Collection

University of Debrecen in Hungary was the location of data collection in the fall of 2021. The spatial skills of 201 and 205 students were measured on paper and using a desktop display, respectively. The paper tests were conducted in a large classroom. The same students used both versions, although four of them were absent from the second measurement. When the desktop display version of the tests was conducted, a computer laboratory was used on multiple occasions as its capacity was 20 people. An LG 20M37A (19.5”) desktop display was used in the case of the latter version. First, the desktop display version was done. The two versions were a few weeks apart so that the students could forget the possible answers. The problems were even randomized. Students came to the tests of their own volition. Naturally, before the tests, the students were informed and verbal consent was obtained from them. Their names were not gathered.

However, some information about the students was asked before each test to help in the evaluation of data. These are the following: age, gender, dominant hand, what they study, and the number of their university attendance years. This information about their characteristics could simply be indicated by drawing an X in the correct place. This procedure was similar among both versions. In the case of the paper version, students had to use a pen to draw an X, while in the case of the desktop display version, they had to press the left mouse button when the cursor was on the appropriate checkboxes.

After a test was done using the desktop display, the results were saved into a CSV file. The results of each participant were written into a column. Each human characteristic was given a separate row. Information about the tests and results were also written into the file: test type, completion time, and the number of correct answers. In the case of the paper version, the results were manually written into an XLSX file. Only the completion time was omitted as it was not measured. In the end, the CSV and XLSX files were merged into an XLSX file.

2.6 Data Analysis

The results were imported into the statistical program package R, and they were evaluated in groups of ten. Before comparing them with each other, the distributions of each dataset were investigated using the Kolmogorov-Smirnov test. This was done to check whether the distributions of the ratios of correct answers are the same or not. We can state that the distributions are significantly different, $p < 2.2\times10^{-16}$.

After that we returned to the comparisons of the expectations of the exact answer rates, measuring the average efficiency of correct fillings. First, all results obtained on paper and display desktop were evaluated and compared to each other. Then, they were grouped and analyzed by test type, and human factors. Afterward, the results that were gathered using the desktop display were evaluated and compared to each other. We have analyzed the data on paper and desktop display separately and compared them as well. The method of evaluation was the application of Welch's test, which is suitable for testing the equality of expectations in case of unequal variances and is a robust method.

3 Results and Discussion

First, the ratio of correct answers itself is investigated on the whole dataset. The positions of the data are illustrated by boxplots in Figure 5.

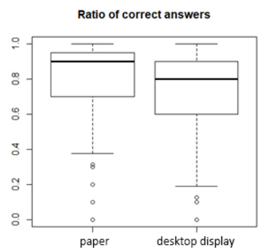

Figure 5
The rates of correct answers on paper and when a desktop display is used

As can be seen in Figure 5, the results on paper are better than those gathered using a desktop display. Comparing the expectations, as shown by Welch's Two Samples test, this difference is significant ($p < 2.2 \times 10^{-16}$). From here onwards, the following abbreviations are used: *M* for mean, *SD* for standard deviation, subscript *P* for tests on paper, and subscript *D* for tests using the desktop display. The values of the means and standard deviations are the following: $M_P = 0.816$, $SD_P = 0.199$, $M_D = 0.739$, $SD_D = 0.216$. Thus, the average ratio of correct answers is 10.41% better on paper and the difference is statistically significant.

To better understand the significant differences, the ratio of correct answers was further analyzed by factors from Section 2.5: test type (3.1), gender (3.2), dominant hand (3.3), age (3.4), and current studies (3.5), with further discussions in Section 3.6. Student attendance years were not analyzed, as most participants were first-year students.

3.1 Ratio of Correct Answers by Test Type

To understand the previously mentioned differences, the rates of correct answers were grouped by test types and the investigation continued on these groups. The rates of correct answers by test types can be observed in Figure 6.

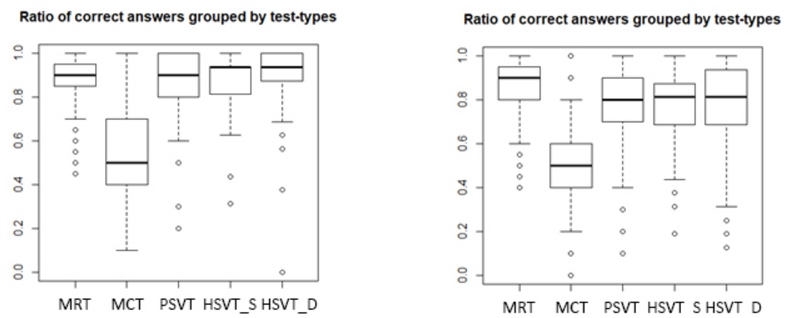

Figure 6

The rates of correct answers grouped by test types and versions: paper version (left), and desktop display version (right)

The ratios of correct answers were compared across both versions. Among the paper tests, all differences were significant except between MRT and HSVT – Synthesis ($p = 0.903$). For the desktop display tests, four comparisons showed no significant differences: MRT vs. HSVT – Decomposition ($p = 0.260$), PSVT vs. HSVT – Synthesis ($p = 0.633$), PSVT vs. HSVT – Decomposition ($p = 0.662$), and between the two HSVT subtypes ($p = 0.887$); all others were significant. Also, the ratio of correct answers is the worst on the MCT type both in case paper and desktop display versions. The lowest correct answer rates were observed for MCT in both formats ($M(MCT)_P = 0.534$, $M(MCT)_D = 0.513$), with slightly better performance on paper. HSVT – Decomposition had the highest paper scores ($M = 0.918$), while MRT was the best on desktop display ($M = 0.873$).

According to Welch's Two Sample t-test, significant differences between paper and desktop versions were found for PSVT ($p < 0.001$), HSVT – Synthesis ($p < 0.001$), and HSVT – Decomposition ($p < 0.001$), with performance increases of 10.99%, 12.35%, and 22.07%, respectively. No significant differences were found for MRT ($p = 0.112$) and MCT ($p = 0.254$), where increases were only 1.83% and 4.09%. Overall, scores were numerically better on paper across all test types, with significant advantages for PSVT and both HSVT subtypes.

3.2 Ratio of Correct Answers by Gender

In the case of paper tests, there were 17 females, whereas 20 females used a desktop display. The remaining participants were males in the case of both versions. According to the results of females $M(female)_P = 0.804$, $SD(female)_P = 0.199$, $M(female)_D = 0.648$, $SD(female)_D = 0.222$, we can conclude that they performed better on paper. A similar phenomenon can be observed when the results of males are investigated: $M(male)_P = 0.817$, $SD(male)_P = 0.199$, $M(male)_D = 0.736$, $SD(male)_D = 0.214$. When comparing the results of females and males on paper, no significant difference could be found among them ($p = 0.557$), while a significant difference exists in the case of the desktop display version ($p = 5.049 \times 10^{-8}$). As can be seen, both males and females performed better on paper, and the difference between the two groups of males is smaller than that of the two groups of females. This can also be observed in Figure 7.

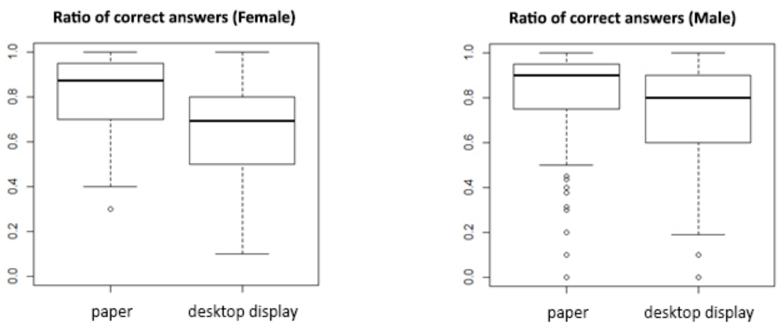

Figure 7
The rates of correct answers grouped by gender: female (left), and male (right)

Even though the difference between the two groups of males is smaller, its significance is stronger due to the larger sample size ($p < 2.2 \times 10^{-16}$). The difference between the two groups of females is also significant ($p < 1.495 \times 10^{-8}$), despite the small sample sizes. However, how did these two genders perform on the tests? Their results are shown in Figure 8.

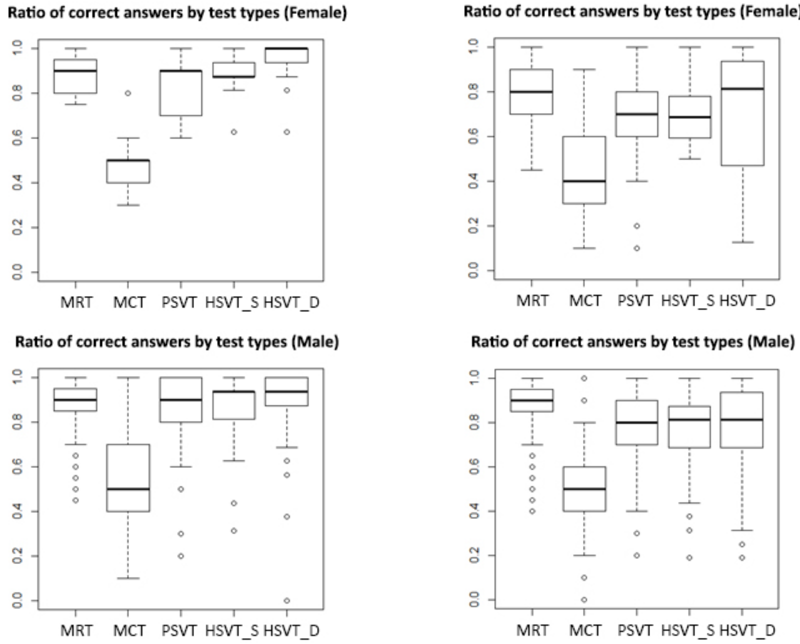

Figure 8

The rates of correct answers grouped by versions: paper (left), and desktop display version (right)

On paper, females showed three insignificant differences between test types: MRT and HSVT – Synthesis ($p = 0.924$), MRT and HSVT – Decomposition ($p = 0.226$), and between the two HSVT subtypes ($p = 0.282$), likely due to the small sample size. Among males, only MRT and HSVT – Synthesis ($p = 0.919$) showed no significant difference, with other comparisons becoming significant due to the larger sample size. When using the desktop display, females had four insignificant differences between their results: MRT and HSVT – Decomposition ($p = 0.260$), PSVT and HSVT – Synthesis ($p = 0.633$), PSVT and HSVT – Decomposition ($p = 0.662$), and between the HSVT subtypes ($p = 0.887$). Males showed three insignificant comparisons: PSVT and HSVT – Synthesis ($p = 0.870$), PSVT and HSVT – Decomposition ($p = 0.315$), and between the HSVT subtypes ($p = 0.341$).

Comparing genders, no significant differences appeared on paper for MRT ($p = 0.812$), MCT ($p = 0.106$), PSVT ($p = 0.281$), HSVT – Synthesis ($p = 0.770$), or HSVT – Decomposition ($p = 0.516$). However, on the desktop display, males outperformed females, with significant differences for MRT (12.77%, $p < 0.001$), MCT (20%, $p = 0.001$), and PSVT (12.84%, $p < 0.001$), and smaller, non-significant differences in the case of HSVT – Synthesis (9.67%, $p = 0.051$) and HSVT – Decomposition (10.67%, $p = 0.218$). This indicates that gender performance differences widen when using desktop displays.

Regardless of immersion level, both genders performed best on the HSVT – Decomposition test (*$M(female)_P = 0.934$*; *$M(male)_P = 0.916$*) and worst on the MCT (*$M(female)_P = 0.482$*; *$M(male)_P = 0.539$*). On the desktop display, both genders again performed best on MRT (*$M(female)_D = 0.783$*; *$M(male)_D = 0.883$*) and worst on MCT (*$M(female)_D = 0.435$*; *$M(male)_D = 0.522$*). Thus, the same test types consistently yielded the highest and lowest scores, regardless of platform.

3.3 Ratio of Correct answers by Dominant Hand

Next, the ratio of correct answers was investigated grouped by the dominant hand of students. On paper, the results of 24 left-handed (LH) and 177 right-handed (RH) were received. Using the desktop display, the spatial skills of 28 left-handed and 177 right-handed students were tested. The results of the former show that $M(LH)_P = 0.843$, $SD(LH)_P = 0.190$, $M(LH)_D = 0.766$, $SD(LH)_D = 0.198$. Contrarily, it is known from the latter that: $M(RH)_P = 0.812$, $SD(RH)_P = 0.200$, $M(RH)_D = 0.721$, $SD(RH)_D = 0.128$. Although left-handed students perform better numerically, no significant differences could be found between left-handed and right-handed students op paper, $p = 0.097$. Using the desktop display however, left-handed students perform significantly better than right-handed ones ($p < 0.001$). The rates of correct answers of both groups are shown in Figure 9.

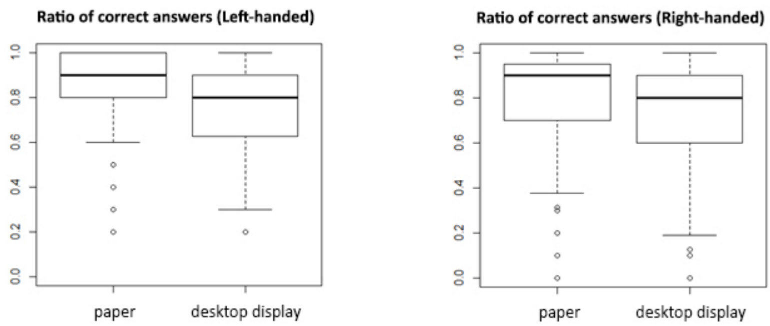

Figure 9

The rates of correct answers grouped by dominant hand: left-handed (left), and right-handed (right)

The two versions were compared in each case. Both groups performed significantly better on paper than when using a desktop display: $p < 0.001$ in the case of both. The increases in their rates of correct answers are numerically 10.14%, and 12.65%, respectively. On paper, left-handed students performed significantly better on the MRT ($p = 0.013$), and HSVT – Decomposition ($p = 0.011$) tests than right-handed ones by 4.29% and 4.38%, respectively. The remaining differences are not significant. Contrarily, when a desktop display is used, three significant differences can be found among them. These can be observed in the case of the MRT ($p = 0.005$), MCT ($p = 0.044$), and PSVT ($p < 0.001$) tests. Therefore, the left-handed students outperformed right-handed ones by 4.14%, 8.67%, and 9.40% on these test types, respectively. While the rates of correct answers in the case of the remaining test types are not significant, left-handed students also performed better on them.

Also, it is possible to achieve better rates of correct answers on paper. In the case of left-handed students, there are two significant differences among them: HSVT – Synthesis ($p < 0.001$), and HSVT – Decomposition ($p < 0.001$). The increases are 18.61%, and 17.94%, respectively. In the case of right-handed students, three significant differences could be found: PSVT ($p < 0.001$), HSVT – Synthesis ($p < 0.001$), and HSVT – Decomposition ($p < 0.001$). Therefore, the rates of correct answers on paper increased by 12.05%, 16.38%, and 17.65%, respectively. Even though the remaining comparisons did not yield significant differences between them, all test types produced better results on paper than with the desktop display in the case of both groups.

Regarding dominant hand, no significant difference can be found on paper ($p = 0.097$). This may be due to the small number of left-handed students. Contrarily, a significant difference can be found among them when a desktop display is used ($p < 0.001$). This means that freshman left-handed students perform significantly better than right-handed ones by 6.18% when using a desktop display. Next, the rates of correct answers were analyzed by test type, and they are shown in Figure 10.

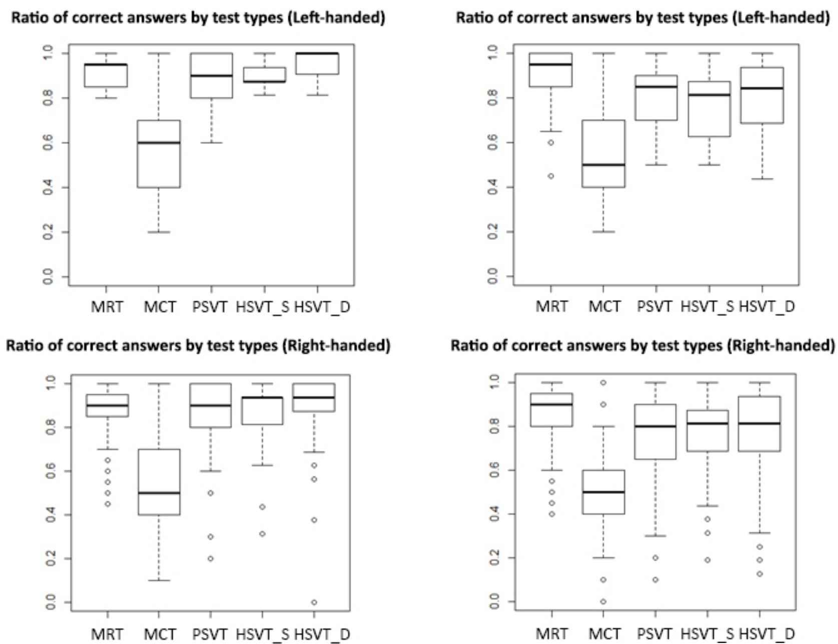

Figure 10

The rates of correct answers grouped by versions: paper (left), and desktop display version (right)

3.4 Ratio of Correct Answers by Age

The following step was to investigate the ratio of correct answers by age. First, the results of age groups are presented in Table 1.

Table 1
The means and standard deviations of the rates of correct answers regarding both versions by age group

Age	N_P	M_P	SD_P	N_D	M_D	SD_D
18	19	0.821	0.192	31	0.712	0.222
19	79	0.808	0.210	80	0.732	0.218
20	61	0.828	0.192	58	0.733	0.207
21	32	0.798	0.189	25	0.703	0.226
22	10	0.856	0.199	11	0.752	0.205

Significant differences exist among the versions in case of each age group. Students performed significantly better on paper than when using a desktop display by 15.30%, 10.38%, 12.96%, 13.51%, and 13.82% in case of the 18, 19, 20, 21, and 22 age groups, respectively. These differences are significant, the respective *p*-values as follows: 5.156×10^{-6}, $5.718\times10^{\wedge-9}$, 1.555×10^{-11}, 4.225×10^{-6}, 0.002. The smallest value of the significance level is the latter, and it is due to the small sample size. Although numerically, the difference is about 10%.

Next, the rates of correct answers were studied on both platforms. First, the rates of correct answers on paper were investigated. These are presented in Table 2.

Table 2
The means of the rates of correct answers of each age group on paper

	Age: 18	Age: 19	Age: 20	Age: 21	Age: 22
MRT	0.884	0.896	0.899	0.853	0.905
MCT	0.584	0.516	0.543	0.528	0.540
PSVT	0.800	0.844	0.867	0.828	0.920
HSVT:S	0.905	0.884	0.901	0.867	0.925
HSVT:D	0.931	0.900	0.929	0.912	0.988

The rates of correct answers of every age group were compared to each other on paper. The number of significant difference is quite small:

- Comparing the ages of 18 and 22, there are two: between PSVT tests ($p = 0.018$), and among HSVT – Decomposition tests ($p = 0.045$);
- Between the ages of 19 and 22, there are also two among the same test types ($p = 0.024$, and $p < 0.001$), respectively;
- Among the ages of 20 and 22, there is only one significant difference, and that is between HSVT – Decomposition tests ($p < 0.001$);

- Lastly, between the ages of 21 and 22, two significant differences could be found: among PSVT tests ($p = 0.019$), and between HSVT – Decomposition tests ($p < 0.001$).

However, it can be observed that these differences only arose when an age group was compared to the age group of 22, and mainly the PSVT and HSVT – Decomposition tests are concerned. The next step was to see whether the same age groups have significant differences using the desktop display version. Regarding that version, the rates of correct answers of students can be seen in Table 3.

Table 3
The means of the rates of correct answers of each age group using a desktop display

	Age: 18	Age: 19	Age: 20	Age: 21	Age: 22
MRT	0.865	0.871	0.876	0.867	0.914
MCT	0.480	0.522	0.536	0.473	0.515
PSVT	0.765	0.773	0.757	0.728	0.821
HSVT:S	0.748	0.766	0.780	0.752	0.739
HSVT:D	0.760	0.789	0.780	0.780	0.784

Similarly as before, the rates of correct answers of every age group using a desktop display were compared to each other. The number of significant difference is also small in the case of this version. According to the results of the comparisons, these differences are the following:

- When comparing the ages of 18 and 20, one significant difference could be found in the case of MCT tests ($p = 0.015$);
- Among the ages of 18 and 22, two significant differences exist: between MRT tests ($p = 0.043$), and among PSVT tests ($p = 0.019$);
- Between the ages of 19 and 22, significant differences can be observed in the case of the same two test types as previously ($p = 0.039$, and $p = 0.024$, respectively);
- Lastly, among the ages of 21 and 22, significant differences occur in the same two test types ($p = 0.046$, and $p = 0.001$, respectively).

Again, students of 22 years of age performed over the other age groups, mainly in the case of MRT and PSVT test types.

The last investigation regarding age groups was to compare the rates of correct answers among the two versions by test type. According to the results of the comparison, the rates of correct answers are always better on paper than when using a desktop display. However, not all differences are significant. Regarding the group of 18-year-old students, only HSVT – Synthesis and HSVT – Decomposition have significant differences among the versions ($p < 0.001$, $p < 0.001$, respectively). In the case of the remaining age groups, significant differences only arose in cases of PSVT, HSVT – Synthesis, and HSVT –

Decomposition tests. Regarding 19-year-old students, the significances are $p = 0.001$, $p < 0.001$, $p < 0.001$ for these tests, respectively. In the case of 20-year-old students, all three $p < 0.001$. Regarding 21-year-old students, the significances of these three tests are $p = 0.005$, $p = 0.013$, $p < 0.001$, respectively. In the case of 22-year-old students, they are $p = 0.004$, $p = 0.004$, $p = 0.002$, respectively.

3.5 Ratio of Correct Answers by Current Studies

The last step was to investigate the rates of correct answers grouped by the current studies of the students. On paper, there were 60 civil engineering (CE), 100 mechanical engineering (ME), and 41 vehicle engineering (VE) students, while when a desktop display was used, their numbers were 59, 104, and 42, respectively. It is known of civil engineering students that $M(CE)_P = 0.814$, $SD(CE)_P = 0.188$, $M(CE)_D = 0.716$, $SD(CE)_D = 0.218$. Contrarily, it is known of mechanical engineering students that $M(ME)_P = 0.817$, $SD(ME)_P = 0.205$, $M(ME)_D = 0.736$, $SD(ME)_D = 0.211$, and of vehicle engineering students that $M(VE)_P = 0.817$, $SD(VE)_P = 0.202$, $M(VE)_D = 0.720$, $SD(VE)_D = 0.224$. The rates of correct answers of all three groups are shown in Figure 11.

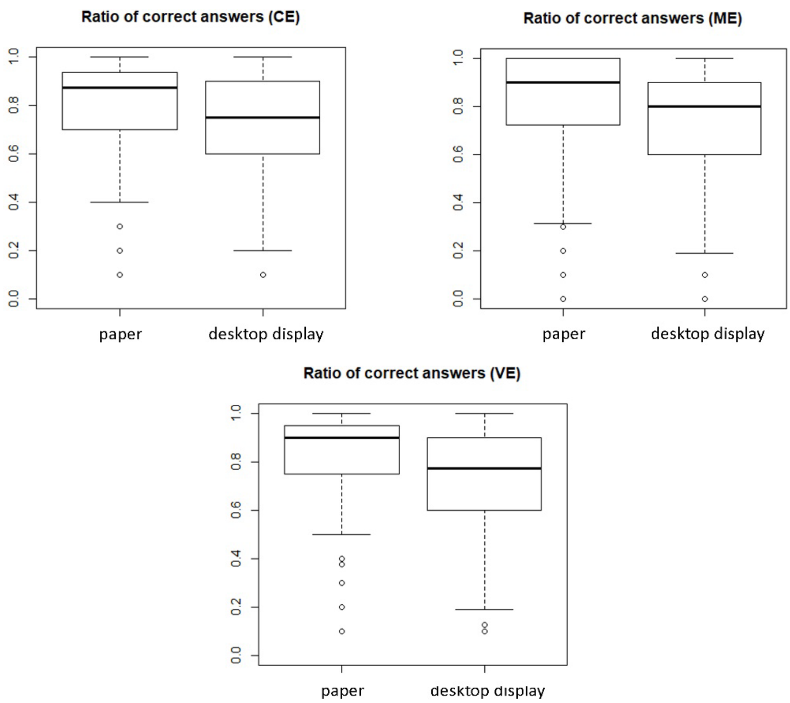

Figure 11
The rates of correct answers grouped by current studies: civil engineering (left), mechanical engineering (right), and vehicle engineering students (middle)

As can be seen in Figure 11, when comparing both versions by studies, the rates of correct answers are better on paper. Significantly, even: $p = 4.363 \times 10^{-12}$ for civil engineering, $p = 7.271 \times 10^{-13}$ for mechanical engineering, and $p = 5.607 \times 10^{-8}$ for vehicle engineering students. The increases in the rates of correct answers are 13.59%, 10.97%, and 13.47%, respectively.

Next, a comparison was done between each test between the versions. The rates of correct answers of students grouped by their current studies are shown in Table 4.

Table 4
The means of the rates of correct answers of each group by current studies on both platforms

	CE_P	ME_P	VE_P	CE_D	ME_D	VE_D
MRT	0.878	0.894	0.894	0.863	0.875	0.886
MCT	0.547	0.534	0.515	0.497	0.530	0.494
PSVT	0.853	0.847	0.844	0.750	0.777	0.753
HSVT:S	0.874	0.893	0.909	0.772	0.768	0.743
HSVT:D	0.917	0.917	0.922	0.780	0.785	0.772

According to the results of the comparison it can be concluded that significant differences could only be found among the PSVT tests and the two HSVT subtests. In the case of civil and mechanical engineering students, all of these significances have the value of $p < 0.001$. Regarding vehicle engineering students, the significance among PSVT tests is $p = 0.001$, while the significances are $p < 0.001$ in the case of both HSVT subtests.

Lastly, all groups of students by studies were compared to each other. On paper, there were no significant differences among the rates of correct answers of the groups. The results are similar in the case of those who used the desktop display: there were also no significant differences among them. Therefore, regarding these three studies, it does not matter which of them did the students choose when they applied to the university. This fact does not influence their spatial skills.

3.6 Further Discussions

It is known from the literature that immersive VR technologies – such as head-mounted displays – can significantly enhance the spatial skills of users as they can receive better results on these tests. This information can be crucial when these tests are taken inside VEs. Since head-mounted displays are expensive, they cannot be always used as display devices to take these tests, especially when the students are high in number. Therefore, the alternative is either to use non-immersive VR technologies – such as desktop displays – or to simply conduct these tests on paper.

According to the results, the latter should be used. As the users cannot immerse themselves in VEs when a desktop display is used, significantly better results can be achieved on paper. This is also true in the case of multiple human characteristics. Naturally, this does not mean that tests should not be taken in non-immersive environments, but the limits of points to pass the tests should be carefully chosen by examination committees.

This paper has some limitations as well. The main limitation of this study is the order of the two measurements regarding the versions of the tests. Although these measurements were done weeks apart and the questions were randomized, a more robust method could involve a counterbalanced design, where half of the students would have completed the paper version first to better control for possible practice effects and task familiarity. This will be addressed in future research and the yielded results could be compared to the results presented in this study. Another limitation is age. Although significant age-related differences were not found consistently, it is possible that older students may have benefited from additional coursework (such as descriptive geometry) not yet taken by younger students. This potential effect of educational exposure could be examined more thoroughly in future work.

Conclusions

After investigating the rates of correct answers of 201 and 205 freshman engineering students in higher education. Each group of users performed significantly better on paper than when using a desktop display. Also, in most cases, the PSVT, HSVT – Synthesis, and HSVT – Decomposition are the types of tests that present significant increases in the rates of correct answers within the two versions. This means that similar results are produced on the MRT, and MCT tests between the desktop display and paper versions. While the paper-based versions yielded significantly higher rates of correct answers, these results are preliminary. Due to methodological limitations, including the lack of a counterbalanced design, further research is required before definitive conclusions about the superiority of one medium over another can be drawn.

Also, since the rates of correct answers of students who were majoring in three different subjects were investigated, it can be concluded that their studies do not affect them in their freshman year. Naturally, spatial skills can be increased in higher education due to the current curriculum of engineering studies. Even though better results can be achieved on paper, improvements can be achieved on desktop display easily due to the digital tests being automatic. Therefore, it would be interesting to see how well would their spatial skills increase, and whether differences would occur among them in the future as they learn different subjects at the university.

The results also show the critical role of technological design in cognitive assessment systems, suggesting that non-immersive VE, while promising, require careful calibration to ensure validity and comparability with traditional testing formats.

Acknowledgement

Funded by the Research Fellowship Programme (Code: 2024-2.1.1-EKÖP) of Ministry of Culture and Innovation of Hungary from the National Fund for Research, Development and Innovation.

This work has been implemented by the TKP2021-NVA-10 project with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the 2021 Thematic Excellence Programme funding scheme.

References

[1] H. Gardner, *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books, 1983

[2] D. F. Lohman, "Spatial Geometry and Spatial Ability-How to Make Solid Geometry Solid," in Selected papers from the Annual Conference of Didactics of Mathematics, P. H. Maier, Ed., 1996, pp. 63-75

[3] A. Yani, A. Mulyadi, and M. Ruhimat, "Contextualization of spatial intelligence: Correlation between spatial intelligence, spatial ability, and geography skills," *J. Balt. Sci. Educ.*, Vol. 17, No. 4, pp. 564-575, 2018, doi: 10.33225/jbse/18.17.564

[4] M. C. Linn and A. C. Petersen, "Emergence and characterization of sex differences in spatial ability: A meta-analysis," *Child Dev.*, Vol. 56, No. 6, p. 1479, 1985, doi: 10.2307/1130467

[5] J. M. Zacks, "Neuroimaging studies of mental rotation: a meta-analysis and review," *J. Cogn. Neurosci.*, Vol. 20, No. 1, pp. 1-19, 2008, doi: 10.1162/jocn.2008.20013

[6] M. Jeannerod and P. Jacob, "Visual cognition: a new look at the two-visual systems model," *Neuropsychologia*, Vol. 43, No. 2, pp. 301-312, 2005, doi: 10.1016/j.neuropsychologia.2004.11.016

[7] M. A. Motes, R. Malach, and M. Kozhevnikov, "Object-Processing Neural Efficiency Differentiates Object from Spatial Visualizers," *Neuroreport*, Vol. 19, pp. 1727-1731, 2008, doi: 10.1097/WNR.0b013e328317f3e2

[8] P. H. Maier, "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, 1996

[9] E. E. Ghiselli, "The validity of aptitude tests in personnel selection," *Pers. Psychol.*, Vol. 26, No. 4, pp. 461-477, 1973, doi: 10.1111/j.1744-6570.1973.tb01150.x

[10] C. Miller, "Enhancing Visual Literacy of Engineering Students through the Use of Real and Computer Generated Models," *Eng. Des. Graph. J.*, Vol. 56, pp. 27-38, 1992

[11] M. Peters, P. Chisholm, and B. Laeng, "Spatial ability, student gender, and academic performance," *J. Eng. Educ.*, Vol. 84, No. 1, pp. 69-73, 1995, doi: 10.1002/j.2168-9830.1995.tb00148.x

[12] A. Nazareth, A. Herrera, and S. M. Pruden, "Explaining sex differences in mental rotation: role of spatial activity experience," *Cogn. Process.*, Vol. 14, No. 2, pp. 201-204, 2013, doi: 10.1007/s10339-013-0542-8

[13] F. D. Glamser and R. W. Turner, "Youth sport participation and associated sex differences on a measure of spatial ability," *Percept. Mot. Skills*, Vol. 81, No. 3, Pt 2, pp. 1099-1105, 1995, doi: 10.2466/pms.1995.81.3f.1099

[14] S. Brownlow and C. A. Miderski, *How Gender and College Chemistry Experience Influence Mental Rotation Ability*. 2001

[15] C. L. Miller and G. R. Bertoline, "Spatial Visualization Research and Theories: Their Importance in the Development of an Engineering and Technical Design Graphics Curriculum Model," *Eng. Des. Graph. J.*, Vol. 55, pp. 5-14, 1991

[16] J. Piaget and B. Inhelder, *Child's Conception of Space: Selected Works*, Vol. 4. Oxon, United Kingdom & New York, NY, USA, 2013

[17] B. J. Orde, *A Correlational Analysis of Drawing Ability and Spatial Ability (Doctoral Dissertation)* 1996

[18] R. Pak, "A further examination of the influence of spatial abilities on computer task performance in younger and older adults," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, Vol. 45, No. 22, pp. 1551-1555, 2001, doi: 10.1177/154193120104502203

[19] J. Rovet, D. R. Olson, and E. Bialystok, "The Education of Spatial Transformations. Spatial Cognition: The Structure and Development of Mental Representations of Spatial Relations," pp. 164-181, 1983

[20] H. H. S. Ip and C. Li, "Virtual reality-based learning environments: Recent developments and ongoing challenges," in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2015, pp. 3-14

[21] C. A. Sanchez, "Enhancing Visuospatial Performance through Video Game Training to Increase Learning in Visuospatial Science Domains," *Psychon. Bull. Rev*, Vol. 19, pp. 58-65, 2012, doi: 10.3758/s13423-011-0177-7

[22] R. N. Shepard and J. Metzler, "Mental rotation of three-dimensional objects," *Science*, Vol. 171, No. 3972, pp. 701-703, 1971, doi: 10.1126/science.171.3972.701

[23] *Special Aptitude Test in Spatial Relations*, Developed by the College Entrance Examination Board. New York, NY, USA: CEEB, 1939

[24] R. B. Guay, "Purdue Spatial Visualization Test - Visualization of Rotations," *Purdue Research Foundation*, 1977

[25] V. Heinrich, *A Spatial Visualization Test for Selecting Engineering Students (Doctoral Dissertation)* 1988

[26] P. Baranyi and Á. Csapó, "Definition and synergies of cognitive infocommunications," *Acta Polytech. Hung*., Vol. 9, pp. 67-83, 2012

[27] G. Sallai, "The cradle of cognitive infocommunications," *Acta Polytech. Hung*., Vol. 9, pp. 171-181, 2012

[28] M. Macik, "Cognitive Aspects of Spatial Orientation," *Acta Polytech. Hung*., Vol. 15, pp. 149-167, 2018

[29] A. Kovari, J. Katona, and C. Costescu, "Quantitative Analysis of Relationship between Visual Attention and Eye-Hand Coordination," *Acta Polytech. Hung.*, Vol. 2020, pp. 77-95, doi: 10.12700/aph.17.2.2020.2.5

[30] H. Mclellan, "Cognitive Issues in Virtual Reality," *J. Vis. Lit.*, Vol. 18, No. 2, pp. 175-199, 1998, doi: 10.1080/23796529.1998.11674538

[31] C. M. Oman, W. L. Shebilske, J. T. Richards, T. C. Tubré, A. C. Beall, and A. Natapoff, "Three dimensional spatial memory and learning in real and virtual environments," *Spat. Cogn. Comput.*, Vol. 2, No. 4, pp. 355-372, 2000, doi: 10.1023/a:1015548105563

[32] T. Guzsvinecz, É. Orbán-Mihálykó, E. Perge, and C. Sik-Lányi, "Analyzing the Spatial Skills of University Students with a Virtual Reality Application Using a Desktop Display and the Gear VR," *Acta Polytech. Hung*., Vol. 17, pp. 35-56, 2020, doi: 10.12700/aph.17.2.2020.2.3