VERSITA

10.2478/jbe-2013-0004

New Ideas in Scoring the Mental Rotation Test

Attila Bölcskei[1], András Zsolt Kovács[2,3] and Domen Kušar[4]

[1]Institute of Architecture,
Ybl Miklós Faculty of Architecture and Civil Engineering, Szent István University, Budapest, Hungary
bolcskei.attila@ybl.szie.hu

[2]Institute of Architecture,
Ybl Miklós Faculty of Architecture and Civil Engineering, Szent István University, Budapest, Hungary
kovacs.andras.zsolt@ybl.szie.hu

[3]Department of Architectural Representation,
Faculty of Architecture, Budapest University of Technology and Economics, Budapest, Hungary
kovacs.andras.zsolt@ybl.szie.hu

[4]Faculty of Architecture, University of Ljubljana, Ljubljana, Slovenia
domen.kusar@fa.uni-lj.si

Abstract: *Spatial ability development is of paramount importance in engineering training, especially for architects. The paper aims to compare results achieved by the world-wide Mental Rotation Test (MRT) at the University of Ljubljana and at the Ybl Faculty of SzIU in Budapest, with respect to total scores, improvement and hand preferences. The paper concludes that the mental rotation aspect of spatial intelligence can significantly be developed by Descriptive Geometry courses in both countries. Sophisticated statistical analysis, however, leads to new ideas in scoring MRT. The main goal of the paper is to present an alternative scoring system, which seems to be fairer and provides the expected statistical behavior of samples.*

Keywords: *spatial ability, Mental Rotation Test, Descriptive Geometry education.*

1. INTRODUCTION

Spatial visualization is a constituent part of spatial intelligence in humans. Together with linguistic, mathematical, locomotory, natural, musical and personal intelligence, it contributes to all human capabilities.

The study of spatial visualization is a relatively young scientific discipline. The first researcher in this field was Sir Francis Galten, who, in 1880, reported about his experimental findings in the field of mental visualization [1].

The idea that spatial visualization can exist separately from general intelligence was an important finding, which was confirmed by the results of psychometric studies. This was later followed by confirmation of the existence of a greater number of components of spatial visualization. However, due to the different analytical techniques used, confusion arose about the different names used for its various constituent parts.

In parallel with research into spatial visualization skills, instruments for its measurement were developed. In 1971 Shepard and Metzler evolved the so-called "Mental rotation test" (referred to in the text which follows as the MRT) [2]. This test will be applied in our researches throughout the paper. Other similar tests also exist, inter alia MCT (the Mental Cutting Test, developed in 1939), DAT (the Differential Aptitude Test, 1990), and TPS (Test Priestorovych Schopnosti - the Spatial Imagination Test, 2003). The DAT test and, particularly, the TPS, aim to cover all of the sub-fields of spatial visualization skills, whereas the MRT is more specialized for the spatial perception of rotations and mirroring, and the MCT for the relationship between the plane view and the actual object. The rapid development of computer software has greatly accelerated the data processing, resulted in new and improved techniques. The MRT was thus subjected to such development, since there was a need to improve it and remove all of the external disturbing elements. Thus, Shiina and Suzuki [3] revised and improved this method. What is more, they unified the degree of difficulty for each studied case.

2. PARALLEL SURVEYS BY MRT IN SLOVENIA AND IN HUNGARY.

Measuring spatial ability at our institutions already has some antecedents. At the University of Ljubljana, Faculty of Architecture, on the Descriptive Geometry course the colleagues have been monitoring the level of spatial perception at the outset of the course already since 1999. For testing, they use the MRT. The experiences of the long lasting research have been collected in the paper of D. Kušar [4]. In Budapest, at the Ybl Miklós Faculty of Architecture and Civil Engineering of SzIU, in the last years, the colleagues conducted some researches on spatial ability of civil engineer and architect students in the frame of Descriptive Geometry courses, too. The results of comprehensive surveys using Mental Cutting Test can be found in [5, 6, 7, 8].

In the academic year 2011/12, the authors decided to perform a parallel measuring by MRT that focuses on architect students in both locations. The similarity of the teaching materials of the Descriptive Geometry courses in Slovenia and in Hungary makes the test-results comparable. In Ljubljana, subjects (= students participating in the survey) were required to complete the test twice in the year, namely on the first and on the last lessons. In Budapest however, where the first semester ends with examination, we measured the development three times: at the beginning of the first semester, and on the last lessons of the first and the second semesters. At the sampling, subjects were identified by their administration code and answered to questions with respect to their gender and hand preference. The total number of subjects was 231 in Ljubljana and 204 in Budapest.

The Mental Rotation Test consists of 20 items. Each item is composed of a criterion figure, two correct alternatives and two distractors. The criterion figure in each item is one of the four different structures used by Shepard and Metzer. Correct alternatives are identical in structure, but shown in different rotated position. The distractors are rotated mirror images (10 items) or rotated images of other structures (10 items). The "Answer Key for MRT" contains the

recommended way of scoring that has been used world-wide. According to this, two points are given if both choices of the subject are correct, none if one choice is correct but the other one is false, or if both are incorrect. If only one design was chosen and it is correct, then one point is given. This scoring system intends to eliminate the need to apply a correction for guessing. Consequently, the total scores are between 0 and 40 points. The subjects have no information about the way of scoring. The time limit for university students is 6 minutes for the whole test.

In the first part of the paper we compare the test results in the two countries from different view-points. Basic statistics, as relative numbers of distribution and indexes will be used in this analysis. In the second part of the discussion we will use hypothesis testing that leads to an alternative scoring system, which seems to be fairer and provides the expected statistical behavior of samples.

2.1. *Comparative survey of test results in Ljubljana and Budapest.*

In the first Table the Reader may find the average scores of the subjects in the two countries. Previous surveys conducted with the Metal Cutting Test revealed [7], that there is a considerable difference between the achievement of males and females. This observation was reconfirmed also in Hungary [5, 6, 7]. Therefore, we expected significant gender difference in the MRT test results, too. The outcomes are collected in Table 1.

Group of students	*September*	*December*	*May*
University of Ljubljana (UL)			
Overall	20.78	-	24.30
Females	17.88	-	22.37
Males	24.65	-	29.61
Ybl Faculty of SzIU (Ybl)			
Overall	20.13	23.95	26.71
Females	17.32	21.13	23.81
Males	22.31	26.15	29.03

Table 1 MRT results with respect to location and gender

Mean scores of the sample represented in the chart of Table 1 show that the values increased in all groups. This simple fact alone, already underlines the importance of Descriptive Geometry in the curricula, namely this is the only subject responsible for the evaluation of spatial imagination in the first semesters.

In accordance with our conjecture, males achieve better results in the MRT than females. This is the case in both countries. The reason of the fact should be studied by psychologists in greater detail.

In Hungary the knowledge of spatial visualization at the entry level is lower than in Slovenia. We explain this effect by the following: in Slovenia the leading training center for architects is in Ljubljana at UL, while in Budapest the best students are accumulated in Budapest University of Technology and Economics. On the other hand, in Ljubljana there is an entrance examination

whose consistent part is a kind of spatial ability test. The research carried out in 2004 proved that this examination is better done by the students with better spatial ability [10].

At the end of the studies the mentioned handicap vanished, the Hungarian students catch up with their Slovenian mates. This again heightens the importance of Descriptive Geometry in the education: it enables the students to make up for misses. The teaching seems to be extremely effective for males in Budapest, where the average development is cca. 7.5 points.

Finally, we remark that the ratio of sexes is basically different in the two locations. In Budapest, it is, roughly speaking, 1:1, but in Ljubljana 3:1 for the women. This explains why the overall mean value is closer to female scores at UL and to the middle at Ybl.

Another interesting aspect of the comparison was the measurement of improvements by introducing relative numbers of distribution. It is not too hard to concede that even the same subject may produce different results within a short time period in the same test influenced by his or her momentary state. According to the scoring system, one single change can produce two points difference in the achievement, therefore we consider improvement or decline at least two point differences between September and May results. Next Figures summarize the results in improvement: the three colors refer to those who improved/worsened their results or stayed stable.

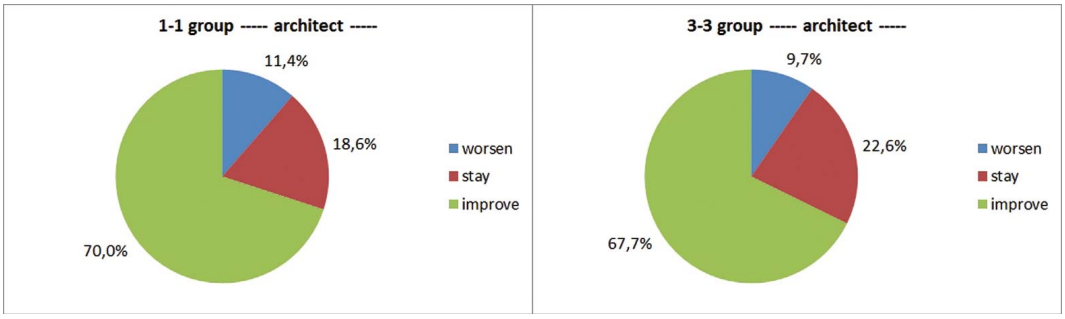

Figure 1 Improvement in whole classes (Budapest: left, Ljubljana: right)

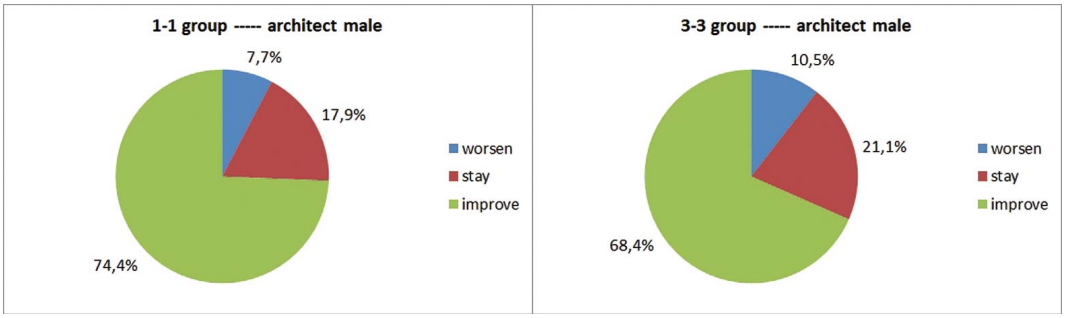

Figure 2 Improvement of male students (Budapest: left, Ljubljana: right)

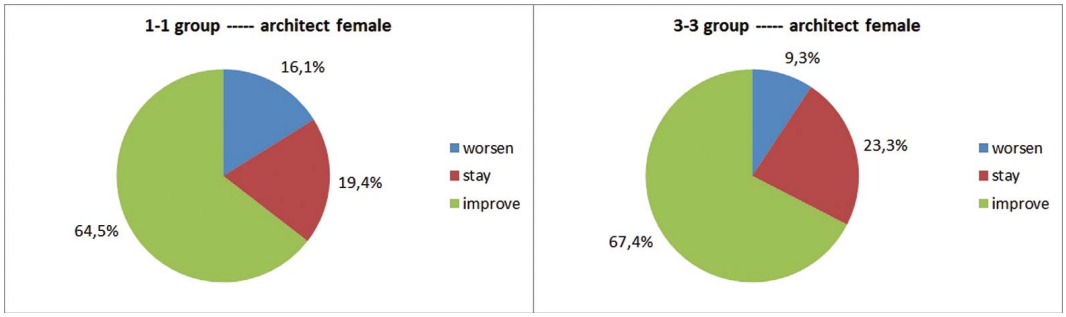

Figure 3 Improvement of female students (Budapest: left, Ljubljana: right)

Discussing the results from this aspect makes it clear that achievements improved in all groups. The Hungarian males performed best: 74% improved and only 8% produced a decline, while the worst rate was due to females: the minimum improvement happened in Budapest by 64,5%, as well as maximum decline by 16%. The rate of constant achievements has its maximum for Slovene females by 23%.

On the other hand, we can conclude that the parallel survey shows analogous tendencies in both countries. There are no significant differences between the Hungarian and Slovenian results if we compare by genders. We can make just two comments. The Hungarian males start their studies from a lower level, but – as we have seen earlier - catch up with the Slovene males – this can stand behind the bigger rate of improvement at Ybl. With respect to females, Ybl students are much less likely to remain constant; the rate of declining part is bigger than in Slovene. The whole classes produce different distribution, due to the definitely bigger rate of females in Ljubljana.

At the end of this subsection we discuss the effect of hand preference to the results of the MRT.

Group of students	***September***	***May***
University of Ljubljana (UL)		
left handed male	26.18 (11 subjects)	16 (1 subjects)
right handed male	24.45	30.23
left handed female	13.89 (9 subjects)	18 (3 subjects)
right handed female	18.17	22.58
Ybl Faculty of SzIU (Ybl)		
left handed male	21.91 (23 subjects)	27.64 (11 subjects)
right handed male	21.30	29.57
left handed female	12.67 (9 subjects)	34 (1 subject)
right handed female	16.74	23.47

Table 2 MRT results with respect to hand preferences

Considering September achievements, left handed males are the best both in Ljubljana, and in Budapest. Their results are about one point better than those of right handed male students. Left handed females, however, produce the worst. What is more, the difference between female students with right and left hand preference is at least four points! The small number of subjects in May makes this kind of comparison impossible.

The only record we can trust is the average score of Slovene left handed males. Their development seems to be dropped behind the right handed males (5.73 vs. 8.27 points). Nevertheless, the obvious diversity in data should initiate further research on the connection of spatial ability and hand preference.

2.2. *Checking normal distribution and its consequence: a new scoring system*

Next we examined whether the distribution of the MRT results could be estimated by a normal distribution at the most commonly used 95% significance level.

We have to mention that we hardly found any hints on the investigation of distribution of tests on spatial abilities in the literature. In [7] the author uses Student's one-sample t-tests, which means that normal distribution was supposed for the Mental Cutting Test results. In [11] [12] and [13] the authors reported on deviation from normal distribution in different cases for different tests. According to our lights the normal distribution of Mental Rotation Test results was not investigated yet.

It is known from statistics literature that the most probable values for the parameters of estimated normal distribution – according to the maximum-likelihood method – are: $m \approx$ sample mean and, $\sigma^2 \approx$ unbiased sample variance. If these values are applied to estimate the parameters of normal distribution, with the discretization of the problem and suitability test the following statistics is calculable:

$$\chi^2 = \sum_{1}^{40} \frac{(g_i - b_i)^2}{b_i}, \qquad (1)$$

where:

g_i – denotes frequency;

b_i – denotes estimated frequency of the corresponding, estimated normal distribution.

The 40 classes of distribution refer to the possible scores (s) i.e. from $0 < s \leq 1$, $1 < s \leq 2$, …, to $39 < s \leq 40$. Considering the number of classes and the number of estimated parameters, we can conclude that the above statistics is a χ^2 distribution with 37 degrees of freedom.

We tried to find answer for the null hypothesis:

$\mathcal{H}_0$: some empirical distribution function = cumulative distribution function of some normal distribution

Interestingly enough, at 95% significance level, this hypothesis is not supportable neither in Ljubljana, nor in Budapest; not for whole classes and not for men or women, with some rare exceptions, as the Reader may find in Table 3.

Group of students (2011/12 survey)	*computed* χ^2 *statistics*	*answer to* $\mathcal{H}_o$ $\left(\text{at } 95\%\ \chi^2 \approx 52.19\right)$
University of Ljubljana (UL)		
overall – September	74.61	NO
men – September	55.48	NO
women - September	68.99	NO
overall – May	54.11	NO
men – May	40.36	YES
women - May	44.66	YES
Ybl Faculty of SzIU (Ybl)		
overall – September	71.23	NO
men – September	73.25	NO
women - September	35.81	YES
overall – December	92.71	NO
men – December	67.11	NO
women - December	77.50	NO
overall – May	111.77	NO
men – May	114.23	NO
women - May	54.38	NO

Table 3 Normality test by χ^2 distribution

The fact that the total classes do not follow the normal distribution is not too much arresting: earlier we already have experienced the differences between genders. The conclusion, however, that even the separate distribution of males or females could not be approximated by the normal distribution is worth investigating further. Figure 4 graphically illustrates the falsehood of normal distribution hypothesis in one case.

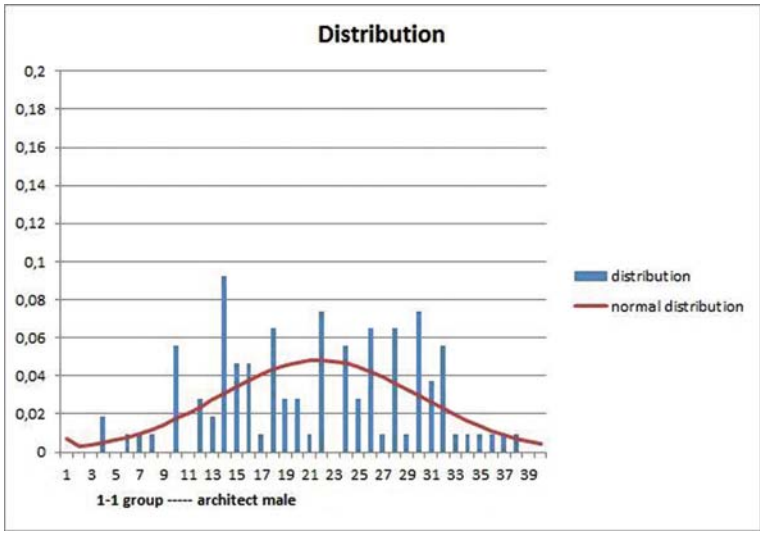

Figure 4 Distribution of males at Ybl in September

In our opinion, considering also the relative large number of subjects, the normal distribution conjecture should necessarily hold. In other words, we think that our samples are representative and the reason why they do not behave adequately lies in the scoring system.

In order to trace the effect of other scoring systems we evidently needed the original test sheets. Unfortunately at University of Ljubljana only the results of the official scoring are available - for the academic year 2011/12. Their one year later survey, however, in 2012 autumn, is already suitable for our purposes, because all the answers are stored. The results of the hypothesis testing of this sample with168 subjects can be found in Table 4.

Group of students (2012/13 survey)	***computed χ^2 statistics***	***answer to $\mathcal{H}_o$ (at 95% $\chi^2 \approx 52.19$)***
University of Ljubljana (UL)		
overall – September	60.74	NO
men – September	59.94	NO
women - September	35.91	YES

Table 4 Normality test by χ^2 distribution in 2012 September at UL

In the first step of our examination we computed the proportion of different type of answers of the subjects. In the joint samples (2011 September at Ybl + 2012 September at UL), these are as follows: 44% of the answers are both good, 25% of them are of type 1 good and 1 bad (mixed); none but two false designs were chosen in 3% of the answers. 4% of the responses are of type only one but good, 1% of them only one but wrong, and in the remaining 23% we have no answers at all (no idea or no time for preparing). Suppose now that a subject randomly fills in the problem sheet – choosing 2 designs in each case. Basic probability theory can prove that in this case cca. 17% of the answers are good-good, 67% are mixed and the remaining are of type false-false. Comparing the results of practice with the theory, it is clear that the two distributions are totally different: more than double of the answers are both correct and less than half of type good-bad what we expected. The number of mixed answers is eight times more than those of two incorrect answers – by random response this ratio would be 4:1. From these we can conclude that the subjects do not guess but try to solve the problems. Right this demonstrates that the mixed answers should somehow be compensated – instead of 0 point score, as we did it until now.

Based on these observations we elaborated a new scoring system that seems to be fairer. The recommended way of scoring is the following:

- 2 points are given if both choices are correct;
- 3/4 point is given for mixed answers;
- 5/4 point is given if only one, but correct design was chosen;
- in all other cases we do not give scores.

Namely, we do think that if someone gives just one answer, than he proves much competence: he knows his limits and is not willing to give an additional answer just by guessing. In our opinion this attitude is worth more than just 1 point for one answer. That is the reason why we offer 5/4 point for them.

On the contrary, if somebody knows just one answer and for the second he/she just guesses – it reflects on less knowledge, that can be requited – for the sake of symmetry – by 3/4 points.

Table 5 and 6 prove that the recommended new scoring eliminates most of the mentioned anomalies and problems.

Group of students	***computed*** χ^2 ***statistics***	***answer to*** $\mathcal{H}_o$ $\left(\text{at } 95\%\ \chi^2 \approx 52.19\right)$
Ybl Faculty of SzIU (Ybl)		
overall – September 2011	35.50	YES
men – September 2011	69.90	NO
women – September 2011	20.42	YES
overall – December 2011	45.82	YES
men – December 2011	43.90	YES
women – December 2011	39.96	YES
overall – May 2012	91.63	NO
men – May 2012	161.76	NO
women – May 2012	37.28	YES

Table 5 Normality test by χ^2 distribution with new scoring system in Budapest

Group of students	***computed*** χ^2 ***statistics***	***answer to*** $\mathcal{H}_o$ $\left(\text{at } 95\%\ \chi^2 \approx 52.19\right)$
University of Ljubljana (UL)		
overall – September 2012	46.50	YES
men – September 2012	47.99	YES
women – September 2012	34.37	YES

Table 6 Normality test by χ^2 distribution with new scoring system in Ljubljana

We have to analyze the unusual behavior of male students in Budapest further. We think that the students have different precognition in spatial thinking that seems to vanish after one semester study of Descriptive Geometry. In Figure 5 we can see the histogram and the estimated normal distribution of males' result in May 2012 at Ybl Faculty.

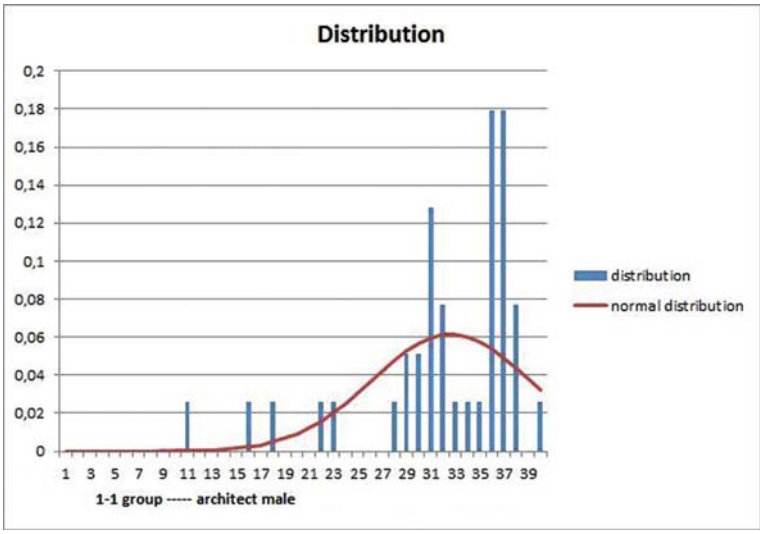

Figure 5 Distribution of males at Ybl in May 2012

We conjecture that the figure refers to the superposition of two or three Gaussian curves. One may belong to the most talented students, whose mean value is at about 37 points; one to the slightly worse ones with the maximum at 31 and probably there is one more of students with little spatial skills. We are planning to investigate this case in a forthcoming publication.

3. CONCLUSION.

In the paper the authors compare the results achieved by the world-wide Mental Rotation Test at the University of Ljubljana and at the Ybl Faculty of SzIU in Budapest. The main observations of the investigation can be formulated as follows:

- Thanks to the analogous structure of secondary and tertiary education systems, there is much correspondence between the results and the observations in Ljubljana and Budapest.
- Descriptive Geometry courses have great impact to the factors of spatial ability measured by the Mental Rotation Test, because both the scores, as the ratio of improving students increase with time.
- There is a considerable difference between the achievements of genders.
- Our conjecture is that left handed females are worse in the MRT than the women with right hand preference. For men, on the contrary, left handed subjects are better, but probably cannot be developed so intensively. All this suggestions must be proved on bigger samples.
- The hypothesis testing of normal distribution led to the revision of the usual scoring of MRT. The authors offer a new way of scoring: two points are given if both choices of the subject are correct, 3/4 if one choice is correct but the other one is false. If only one design was chosen and it is correct, then we recommend 5/4 point for it. This system seems to be fairer and provided the expected statistical behavior of samples in most cases.

We intend to continue our work on analyzing the new score system compared to the old one. Since the new system has an impact to the average score, it is impossible to compare new results with old ones. Thanks to information technics we can simply use both systems parallelly in the future.

REFERENCES

[1] **Mohler, j.,** A review of spatial ability research. *Engineering Design Graphics Journal* **72/2** (2008), *19-30.*

[2] **Shepard, R. N., Metzer, J.,** Mental rotation of three-dimensional objects. *Science* **171**, (1971), *701-703.*

[3] **Shiina, K., Suzuki, K.,** Design of modified mental rotation test and its error analysis. *J. Geometry and Graphics* **3/2** (1999), *211-219*.

[4] **Kušar, D.,** Oscillating conceptions of space of architecture students (in Slovene). *AR, Arhitektura raziskave, Architecture Research* **1**, (2010), *46-51.*

[5] **Németh B., Hoffmann M.:** Gender differences in spatial visualization among engineering students. *Annales Mathematicae et Informaticae* **33** (2006), 169–175.

[6] **Németh B., Sörös Cs., Hoffmann M.:** Typical mistakes in Mental Cutting Test and their consequences in gender differences. *Teaching Mathematics and Computer Science* **5/2** (2007), 385–392.

[7] **Németh B.:** Measurement of the development of spatial ability by Mental Cutting Test. *Annales Mathematicae et Informaticae* **34** (2007), 123–128.

[8] **Bölcskei, A., Gál-Kállay, Sz., Kovács, A. Zs., Sörös, Cs.,** Development of Spatial Abilities of Architect and Civil Engineer Students in the Light of the Mental Cutting Test, *Journal for Geometry and Graphics,* Volume **16/1** (2012), *97–109.*

[9] **Tsutsumi N.**: Evaluation of Spatial Abilities by a Mental Cutting Test at Musashino Art University [Japanese]. Proc. 1990 Annual Meeting of Japan Society for Graphic Science, XIII (1990).

[10] **Kušar, D.,** Spatial ability of students from the Faculty of architecture in Ljubljana (in Slovene). *AR, Arhitektura raziskave, Architecture Research* **1**, (2004), *66-69.*

[11] **Górska, R., Juščáková, Z.,** A Pilot Study of a New Testing Method for Spatial Abilities Evaluation. *Journal for Geometry and Graphics,* Volume **7** (2003), *237–246.*

[12] **Hegarty, M., Waller, D.,** A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence* **32/2**, (2004), *175-191.*

[13] **Maeda, Y., Yoon, S.Y.,** A Meta-Analysis on Gender Differences in Mental Rotation AbilityMeasured by the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT:R). *Educ Psychol Rev* **25**, (2013), *69-94.*