Journal for Geometry and Graphics
Volume 13 (2009), No. 2, 231–240.

Qualitative Analysis of Subtest1 of the TPS

Zuzana Juščáková[1], **Renata A. Górska**[2]

[1] *National Institute for Certified Educational Measurements NÚCEM*
P.O. Box 26, SK 830 00 Bratislava, Slovak Republic
email: zuzana.juscakova@nucem.sk

[2] *Faculty of Architecture, A-6, Cracow University of Technology*
Warszawska st. 24, PL 31-155 Kraków, Poland
e-mail: rgorska@pk.edu.pl

Abstract. The goal of this part of research is to provide a statistical analysis of the first part of the TPS test (the Spatial Imagination Test) developed by Zuzana Juščáková, namely that of Subtest1. The items of Subtest1 have been analyzed in context of their difficulty. An analysis of distracters has been provided, and the correlation between the task items of the test TPS and the standardized psychometric tests ISA and OTRS has been calculated.

Key Words: Classical testing theory, item analysis, educational measurement, test development, spatial ability

MSC 2000: 51N05

1. Introduction

Spatial ability is the trait which may not be measured directly. However, it can be evaluated indirectly by the examination of the skill-levels of a person performing specific tasks. The objective of the spatial test construction is to develop a modern tool, which might be used for spatial ability evaluation both among the students studying technical contents at their studies and also among the practicing engineers who already have started their vocational life.

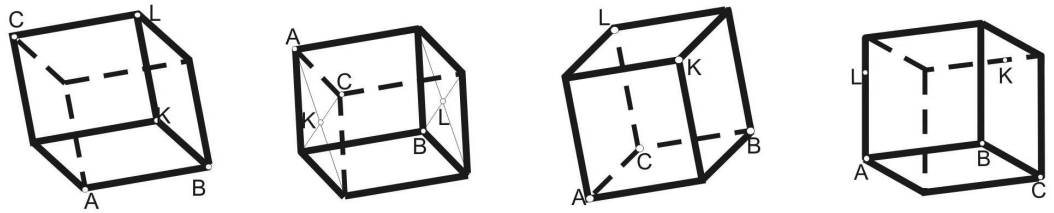
As a new testing instrument, TPS (*"Test Priestorovej Schopnosti"*, which can be translated into English as *"Spatial Imagination Test"*) has been developed by Zuzana Juščáková ([6, 7, 8]). The test itself consists as a whole of three parts: Subtest1, Subtest2 and Subtest3, which have been described in detail in previous publications. TPS has been modified and reconstructed over several years of research and testing ([3, 4, 5]). The test TPS has been administered on a representative group of students ($n > 1450$ subjects) at various types and levels of schools in the Slovak and the Czech Republic, in Poland and in Austria. An experimental subset of the data has been chosen to undergo a further statistical evaluation. The objective of the research work presented in this paper is to provide the results of a qualitative analysis of the first part of the test TPS, which is called Subtest1.

ISSN 1433-8157/$ 2.50

2. Research methodology description

2.1. Subtest1 – problem example

Two different versions of Subtest1, each including 10 items, have been constructed. These two versions will be hereinafter called TPS1A and TPS1B, respectively. In Subtest1 the subject has to identify the mutual position between the line KL and the plane ABC (Fig. 1). The points K, L and A, B, C, which respectively span either the line KL or the plane ABC, have been specified on the faces or edges of a regular polyhedron. The polyhedron can be a cube, tetrahedron or octahedron. The task is to identify this single item out of four possibilities, in which the relation between the line KL and the plane ABC is different from the relations existing in the other options. To give an example, we can see in Fig. 1 that the solution is b. Solution time provided for Subtest1 is 13 minutes. This part of the test has been designed to measure the knowledge of figural relations.



(a) (b) (c) (d)

Figure 1: Sample problem of Subtest1: Parallel or intersecting?

2.2. Subtest1 – research

The standardization of the test, its reliability, differentiability, and the test's validity have been analyzed and the results presented in earlier publications ([4, 5]). In the current evaluation a Cronbach's alpha testing has been performed to evaluate the reliability of the test. The further part of this research focuses on the psychometric quality analysis of each of the test's items. An individual item analysis aims at the specification of their individual difficulty in relation to the various subgroups of subjects solving these items. We provide an analysis of the distracters, we calculate the correlation between particular items and finally we analyze the quantity and quality of the items, which have not been solved due to their difficulty or because they have not been reached within the set time-limit. The first mentioned items will be called "skipped items", the latter "non-reached". The term "skipped item" has been used in such a case, when the testing person (hereinafter called "the subject") moved to a subsequent test item without solving a problem at an earlier position. Further on, we provide a correlation analysis between the results obtained by measurement provided with the TPS1 and with the use of the standardized psychometric tests ISA [1] and OTRS [10]. We also calculate the correlation of these results with individual grades in mathematics and descriptive geometry.

2.3. Group demography

The test has been administered both at high schools and at various types of universities in the above mentioned states. The subset of 1450 subjects has been taken into consideration to undergo a statistical analysis. The structure and proportions of various types of schools participating in the research group have been presented in Fig. 2, where *"sps"* and *"sou"*

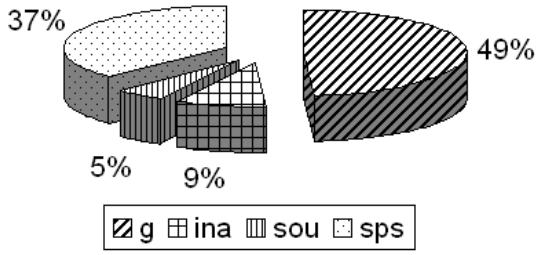


Figure 2: Demographics distribution

denote technical schools, *"g"* is used for a general type of a high school and *"ina"* denotes other types of schools, such as music, art or economic school. The data have been obtained in international cooperation in Slovak and Czech Republics, Poland and Austria.

From the whole experimental group a representative subset of results, that is a limited number of cases, has been picked out for further analysis. Various factors were taken into account in order to ensure sample group representativeness. The grades in mathematics and descriptive geometry represent the levels of success at school. The data was analyzed for the relevant subsets in each group. For TPS1A (Form A of Subtest1) a sample of $n = 283$ subjects and for TPS1B (Form B of Subtest1) a sample of $n = 292$ subjects has been picked out. Statistical values of the mean, median, mode and standard deviation for the test results were compared between the sample and the whole group (see Table 1). The mean scores have been calculated in absolute values. The maximum score for each individual subtest of TPS Subtest1 is 10 points (absolute value), while for the whole TPS test it is 30 points.

Table 1: Comparison of the whole statistical group with the experimental sample

Whole tested group				*Experimental group*							
Mean	Median	Mode	St.dev	Mean	Median	Mode	St.dev	Mean	Median	Mode	St.dev
TPS1 ($n = 1450$)				**TPS1A** ($n = 283$)				**TPS1B** ($n = 292$)			
3,92	4	2	2,43	3,97	4	3	2,38	3,57	3	2	2,27
ISA ($n = 971$)				**ISA** ($n = 283$)				**ISA** ($n = 289$)			
5,81	6	6	3,28	5,98	6	5	2,98	5,81	6	7	3,19
OTRS ($n = 973$)				**OTRS** ($n = 283$)				**OTRS** ($n = 289$)			
12,02	13	15	4,33	12,21	12	12	3,97	11,62	12	11	4,49

3. Analysis of Subtest1

Figure 3 shows the histograms of the results in percentage values for TPS1A and TPS1B measured in the sample groups, while Fig. 4 shows the relative histogram for which the results were taken from the whole sample ($n = 1450$) of the researched group and measured with both forms of TPS1, namely A and B. The graphs characterize a normal distribution. In Table 2 the psychometric characteristics for both versions of the test, that is for TPS1A and

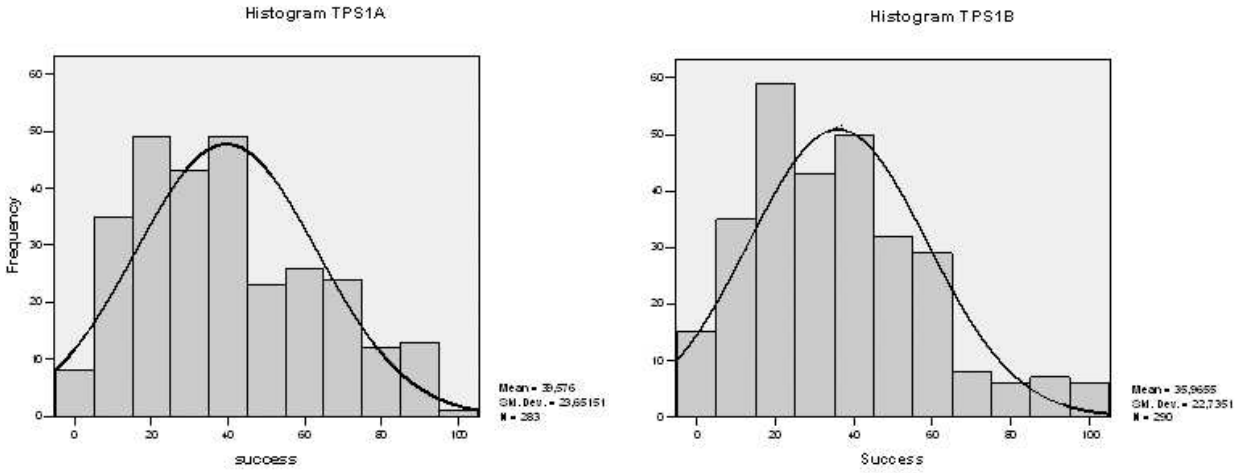


Figure 3: Normal distribution of the results on TPS1A and TPS1B

TPS1B have been presented separately. The data was calculated for the experimental sample of subjects. The mean score, standard deviation and the other parameters were calculated in percentage values. The *Cronbach's alpha value* was respectively equal to 0,65 and 0,64 for TPS1A and TPS1B. According to [9] internal consistency of the test is usually measured with Cronbach's alpha, a statistic calculated from the pairwise correlations between the items. The internal consistency ranges between zero and one. A commonly accepted rule of thumb is that an α of 0,6 – 0,7 indicates acceptable reliability.

3.1. Items' analysis

All the items of Subtest1 have been thoroughly analyzed one by one. As it is difficult to provide all the results here so let us focus on the problem Item 1 from the TPS1A and TPS1B forms (see Fig. 1). We examine the numbers of subjects who skipped or did not reach the certain problem item in the required time limit (see exemplary values provided in Table 4

Table 2: Psychometric characteristics of TPS1A and TPS1B

Subtest	No of cases	Mean[%]	Std.dev.[%]	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
TPS1A	283	39,6	23,7	0,0	85,9	1,4	36,8	42,3	0,65	13,9	27,3
TPS1B	290	36,0	22,7	0,0	80,5	1,3	33,3	38,6	0,64	13,6	26,6

where:

[1] Intervals estimate of success for population – lower limit
[2] Intervals estimate of success for population – upper limit
[3] Standard error of means success
[4] 99% confidence interval - lower limit
[5] 99% confidence interval – upper limit
[6] Cronbach's alpha
[7] Standard error of measurement for success
[8] Intervals estimate of success for individual student

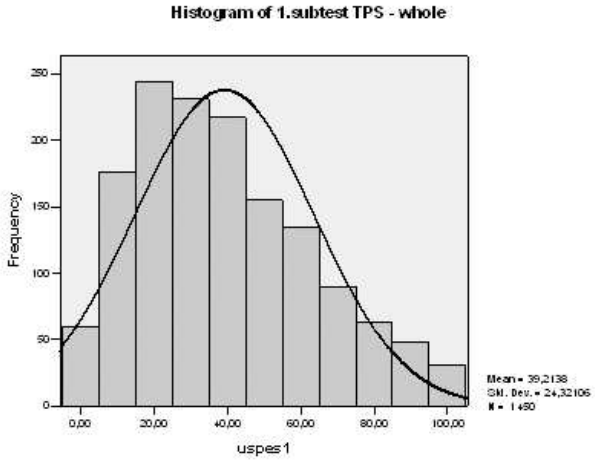


Figure 4: Normal distribution of the results on TPS1 in a whole group

for Item 1). The analysis will be conducted for all the items of Subtest1 in both forms A and B. We calculate the point biserial for evaluation of the general item correlation for each of the suggested solutions a, b, c, d (Table 3). The *point biserial correlation coefficient* (rpb) is a correlation coefficient used when one of the variables is dichotomous (bipartition of the elements). To give an example, let us analyze Item 1 of TPS1A in detail. It can be observed that for the correct answer b the point biserial value 0,38% has the highest value with 58% of the researched population that has chosen this answer to be correct.

Table 3: Point biserial calculation for the Item 1 of TPS1A

Items	*a* 01	*b* 01	*c* 01	*d* 01	*x* 01
Point biserial	-0,18	**0,38**	-,18	-,017	-,009
p[%]	0,13	**0,58**	0,23	0,06	0,00
N	38	**163**	64	16	1

Simultaneously, the analysis of the results' distribution has been provided. The research group of subjects (number of cases $n = 282$ for TPS1A) has been split into 10 subgroups. This part of analysis may be called "the distribution of success". It analyzes the relation between the whole test result and the result reached on a certain item by the subgroup. The x-axis in Fig. 5 represents 10 subgroups of students: the left side of the graph represents the groups of students who received lower scores on the test, while on the right side the scores rise to maximum. The attached graph represents results for the test Item 1. The diagonal of the chart would approximate an ideal graph.

Table 4: Item 01 of the TPS1A

Difficulty	*Skipped*	*Non-reached*	*Point biserial*
42,4	0,00	0,35	0,38

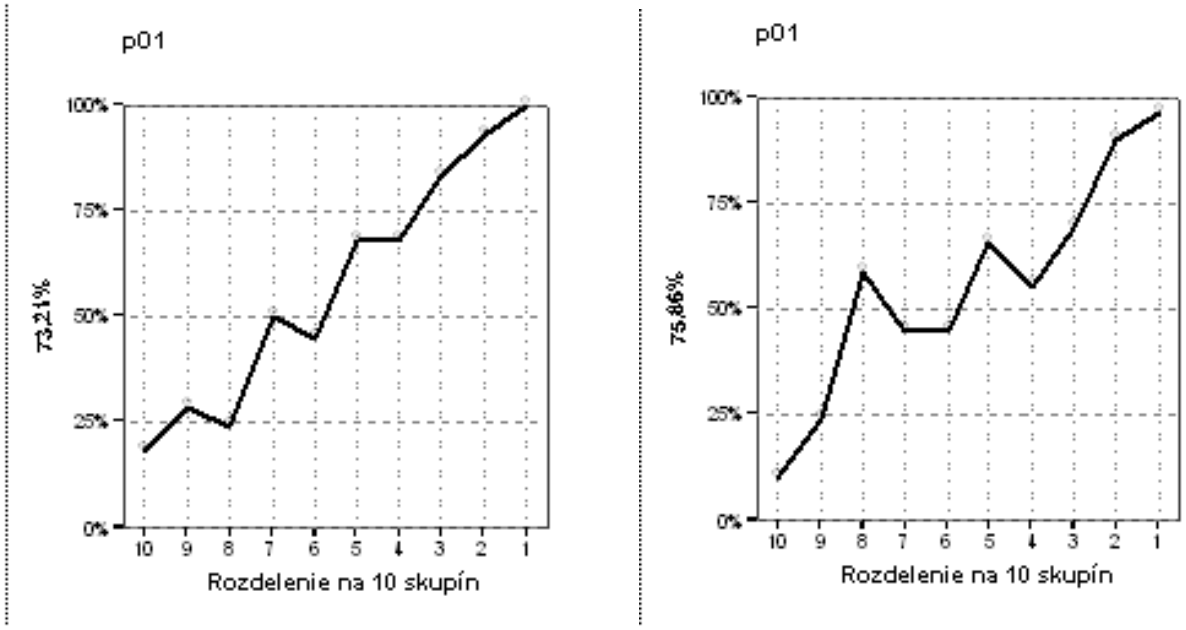


Figure 5: Distribution of success on TPS1A and TPS1B: Item 1

3.2. Items' difficulty

Each item of the form TPS1A (and TPS1B) has been analyzed in context of its difficulty by means of the percentage value of incorrect answers. The higher the percentage value of incorrect answers the item occurred to be more difficult. The results are presented in Table 5. For the version A of the TPS the items 6, 7, 9 and 10 were more difficult for the subjects, while the items 1, 2 and 5 were less difficult. The overall "difficulty" of the test has been presented graphically in Figure 6, where x-axis represents the particular numbers of problem items. The data has been ordered in diminishing range.

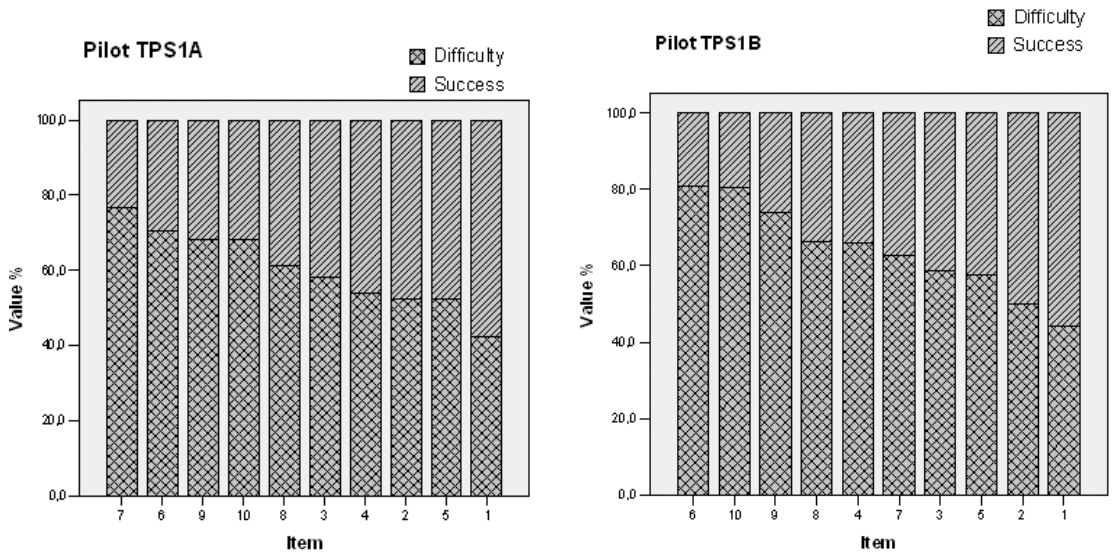


Figure 6: Distribution of success on the TPS1A and TPS1B

Table 5: Difficulty analysis for TPS1A

Item No	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Difficulty	42,4	52,3	58,3	54,1	52,3	70,7	76,7	61,1	68,2	68,2
Std.err.	2,9	3,0	2,9	3,0	3,0	2,7	2,5	2,9	2,8	2,8
Skipped items	0,0	6,36	2,47	1,77	2,47	3,18	2,83	4,24	0,71	4,95
Non-reached items	0,35	0,35	0,35	0,35	2,83	3,89	7,42	12,72	19,08	19,08
Non-solved items	0,35	6,71	2,83	2,12	5,30	7,07	10,25	16,96	19,79	24,03

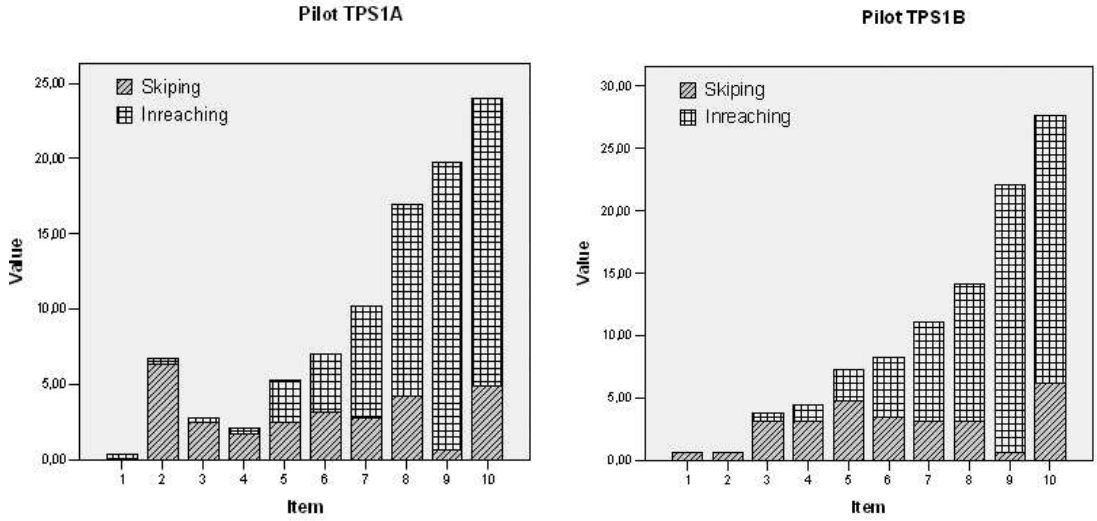


Figure 7: TPS1A and TPS1B: skipped and "non-reached" items

3.3. Non-solved items

Particular items have been also analyzed in context of being not solved. Some problems were not reached by the subjects within a due time. There were also some problems skipped by the subjects due to their difficulty. These cases will be called respectively "non-reached" and "skipped" problems. Table 5 presents the percentage values of the analysis. It is interesting that Item 2 was skipped in many cases (6,36 %). This item will probably be re-designed. Item 10 has not been solved in 24,03% of cases. The question arises what happens in case the items are re-ordered; what would be the proportions of non-solved or skipped items? For TPS1B the problem number 10 has been skipped in 6,21% and non-reached in 21,36% cases. Problem number 5 of this part shows also high value for skipping (4,83%). This aspect of the research will undergo further investigation.

3.4. Distracters' analysis

Distracter is one of the incorrect answers presented as a choice in the multiple-choice test. In this part of analysis we will examine the proportions between correct and incorrect answers given by the subjects who presented various levels of test-efficiency. Such approach will be helpful to determine various aspects of the items' construction and evaluation. We will try

to estimate the following parameters for each item: its discriminating power, its uniformity, self-coherency and its uniqueness. We will investigate the formulation of the problem.

The evaluation of the distracters was based on the following criteria:

- large number of subjects who have chosen the correct variant out of the distracters (according to the answer-key);
- positive value of point biserial: greater than 0,20 for the correct variant;
- negative point biserial value for the incorrect answer, i.e., when the distracter was chosen.

If any of the above criteria is not fulfilled then the item does not follow the criteria of correct construction. Table 3 presents values for Item 1 of the test TPS1A. The correct answer has been indicated by bold face. The distribution of the results for almost all items is similar to that of Item 1. Items 7 and 10 present diversified values (see Table 6). These items were the most difficult for the subjects.

Table 6: Point biserial calculation for items 4, 7 and 10 of TPS1A

Item 4	$a\,04$	$b\,04$	$c\,04$	$d\,04$	$x\,04$
Point biserial	-0,09	-0,17	**0,21**	**0,06**	-0,15
p[%]	0,18	0,16	**0,34**	0,28	0,04
N	51	46	**99**	81	13
Item 7	$a\,07$	$b\,07$	$c\,07$	$d\,07$	$x\,07$
Point biserial	**0,01**	-0,04	**0,19**	-0,03	-0,19
p[%]	0,22	0,28	**0,23**	0,17	0,10
N	62	78	**66**	48	29
Item 10	$a\,10$	$b\,10$	$c\,10$	$d\,10$	$x\,10$
Point biserial	**0,14**	-0,08	**0,00**	**0,02**	-0,07
p[%]	0,15	0,12	**0,32**	0,17	0,24
N	43	35	**90**	47	68

3.5. Correlation with the test items

The reliability of the test will be ensured when the items are homogeneous, i.e., they present internal consistency. Internal consistency of the test means that particular items correlate with each other as they measure the same psychometric value. The correlation between the levels of success on particular item and the overall success on the whole item will be measured by the point biserial value. The boundary values can be distinguished for items which do not correctly rank the subjects into relevant performance categories. It means that the subjects with good performance level on the whole test solved the specific item incorrectly and vice versa, the subjects showing low performance on the test as a whole were successful on the specific item. In case the point biserial value is close to 0, then the item does not categorize the subjects properly. The item with point biserial value greater than 0,25 is considered to be correctly designed as a pedagogical tests and has a good measure. Figure 8 shows graphically calculated values for TPS1A and TPS1B. Items 3, 7 and 10 in TPS1A and Items 9 and 4 in TPS1 B have low values of point biserial.

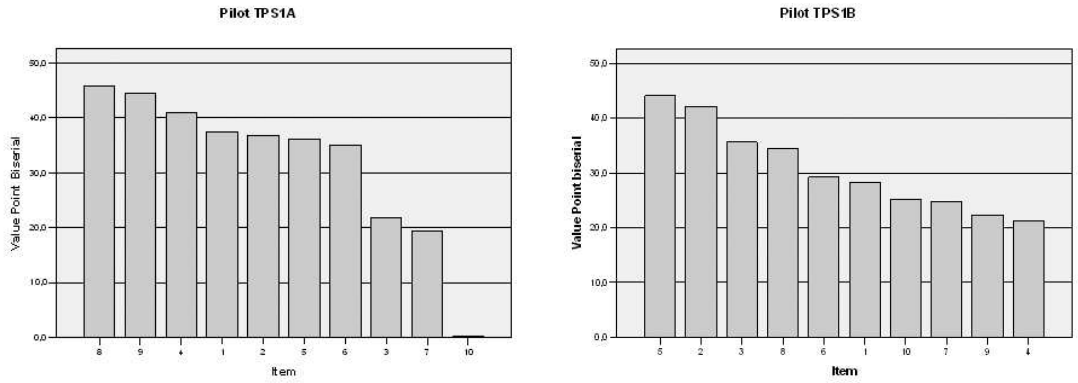


Figure 8: Point biserial values on particular items for TPS1A (*left*) and TPS1B (*right*)

Table 7: Pearson's product correlation coefficient between TPS1A, TPS1B, ISA, and OTRS

	ISA	OTRS
TPS1A	$r = 0,373$ $p = 0,000$ $(n = 283)$	$r = 0,334$ $p = 0,000$ $(n = 283)$
TPS1B	$r = 0,367$ $p = 0,000$ $(n = 289)$	$r = 0,237$ $p = 0,000$ $(n = 289)$

4. Correlation with the ISA and OTRS tests

In order to provide a reference measure of the TPS test, the other standardized tests such as ISA [1] and OTRS [10] have been administered in the researched groups. In Table 7 we present the Pearson's product correlation coefficient r between particular tests. The significance p of the results has been calculated with the use of 2-tailed student's t-test. As the correlation is significant at the 0,01 level, we can observe in all calculations that $p = 0,000$. In Table 8 we can see the correlations between TPS1A (and TPS1B) and grades in mathematics or descriptive geometry. TPS1A shows a good correlation with the performance in the descriptive geometry course, while this can not be observed on TPS1B. There exists good correlation between the TPS1B and the grades in mathematics.

5. Conclusions

The Subtest1 of the TPS test has been thoroughly analyzed. The mean score obtained on TPS1A was 39,7% and on the TPS1B 36,0%. The standard deviation was respectively equal to 23,7% and 22,7%. The optimal values for the test "difficulty" should range between the minimum 20% and the maximum 80%. In TPS1A no item dropped out of this range. The values of point biserial should be in range 0,20 – 1 (see Fig. 8). Only one item of TPS1A (number 10) has a point biserial value of 0,001. The subjects simply did not reach this item

Table 8: Pearson's product correlation coefficient between TPS1A, TPS1B, mathematics and descriptive geometry

	Grade in Math	Grade in DG
TPS1A	$r = -0,074$ $p = 0,228$ $(n = 270)$	$r = -0,377$ $p = 0,0000$ $(n = 134)$
TPS1B	$r = -0,226$ $p = 0,0000$ $(n = 275)$	$r = -0,201$ $p = 0,052$ $(n = 292)$

within a time-limit assigned for solving the test. In TPS1A Item 2 and in TPS1B Item 5 show a high level of skipping. These items should be thoroughly analyzed. The TPS test was standardized on a representative group of subjects ($n > 1450$) and gave fairly good correlations with typical intelligence tests such as ISA and OTRS.

References

[1] E. Fay, G. Trost, G. Gittler: *Intelligenz Struktur Analyse*. Swets and Zeitlinger B. V., Swets Test Services, Frankfurt, Psychodiagnostika a.s., Bratislava 2001.

[2] J. Ferjenčik: *Úvod do Metodológie Psychologického Výskumu*. Portal, Prague 2000.

[3] R. Górska, Z. Juščáková: *A Pilot Study of a New testing Method for Spatial Abilities Evaluation*. J. Geometry Graphics **7**, no. 2, 237–246 (2003).

[4] R. Górska, Z. Juščáková: *Spatial Abilities Test TPS – Further Development and New Results*. Proc. ICEE'05, Gliwice 2005, pp. 76–80.

[5] R. Górska, Z. Juščáková: *TPS test Development and Application into Research on Spatial Abilities*. J. Geometry Graphics **11**, no. 2, 223–236 (2007).

[6] Z. Juščáková: *Mentálne Reprezentácie Priestorovej Predstavivosti*. Proc. 21st Conference on Geometry and Computer Graphics, LDF MZLU, Brno 2001, pp. 48–51.

[7] Z. Juščáková: *Tvorba Testu Priestorovej Predstavivosti II*. Proc. 12th Conference Moderni Matematické Metody v Inženýrství, VŠB Ostrava 2003, pp. 67–70.

[8] Z. Juščáková: *Tvorba Testu Priestorovej Predstavivosti III*. Proc. 14th Conference Moderni Matematické Metody v Inženýrství, VŠB, Ostrava 2005, pp. 56–60.

[9] W. Mendenhall: *Introduction to Probability and Statistics*: Duxbury Press, Ma 1979.

[10] J. Vonkomer: *Orientačný Test Rozumových Schopnosti*. Psychodiagnostické a Didatické Testy, Bratislava 1981.

[11] M. Zvariková: *Measuring Spatial Imagination by Means of Psychological Test*. Proc. CO-MAT-TECH, MTF STU Trnava 2004, pp. 1545–1549.

[12] M. Zvariková, Z. Juščáková: *Gender Differences in Tests of Space Abilities*. Proc. 24^{th} Conference Geometry and Computer Graphics, ČVUT Praha 2003, pp. 313–317.

Received August 20, 2008; final form December 1, 2009