Journal for Geometry and Graphics
Volume 19 (2015), No. 1, 133–157.

How to Develop Spatial Ability? Factors, Strategies, and Gender Specific Findings

Günter Maresch

University of Salzburg
Hellbrunnerstr. 34, A-5020 Salzburg, Austria
email: guenter.maresch@sbg.ac.at

Abstract. The article discusses the results of the analyses of the spatial ability project GeodiKon. 903 students took part in this project in which the major aim was to find out whether the training of each factor of spatial ability and of a repertoire of strategies for solving spatial tasks will lead to an improvement of spatial ability. It will refer to the findings regarding the use of different strategies for solving spatial tasks, promising strategies for solving spatial tasks, and gender specific results.

Key Words: spatial ability, strategies, gender-sensitivity, visualization, spatial relations, mental rotation, spatial orientation, holistic strategy, analytical strategy, spatial thinking, planar thinking, move self, move object, falsifying strategy, verifying strategy.

MSC 2010: 97G80, 51N05, 97G80

1. Introduction

Starting from the beginning of the twentieth century researchers have increasingly been convinced that intelligence is not just one-dimensional. They identified differing aspects of intelligence and defined it as a multi-dimensional term. Ever since there has been clear evidence that spatial ability is one of the fundamental parts of human intelligence [10, 48, 54]. In a further step investigations were made to find the constituing parts of spatial ability. Research work centred on the definition of structured, psychometric, and factor based models of spatial ability [8, 10, 12, 18, 35, 55, 56]. Continuing research made it evident that spatial ability is not defined only by genetical inheritance and is therefore not limited. Individuals can improve their spatial ability actively trough specific support and well balanced training [15, 16]. This awareness is one of the reasons why it is a central aim of geometry lessons to support and train spatial ability of the students. The teachers' preparation for each lesson is based on the curriculum for the subject. In addition to this the competence model for "Geometrisches Zeichnen (GZ)" (- this subject can be described as the subject "Descriptive Geometry" for

ISSN 1433-8157/$ 2.50

lower secondary schools -) [41], the competence model of “Darstellende Geometrie (DG)” (Descriptive Geometry for upper secondary schools) [30], the competence model for mathematics (4^{th} and 8^{th} school level: [4]; 12^{th} school level: [31]), and often school books are relevant for the preparation of lessons. Does also the psychological knowledge about spatial intelligence and the factors of spatial ability — especially because of the fact that we can actively support and train and improve spatial ability of students — play an important role for the didactical design of teaching? The analysis of geometric tasks from mathematics lessons of the first up to the tenth school level shows that mainly the factor visualization is addressed. The other factors are not affected or play just a subordinate role ([35, p. 237]). This situation raises the question whether it is possible to train and improve spatial ability of individuals if the tasks of geometry lessons are selected with the focus on addressing all the factors of spatial ability in a well-balanced proportion? There is a clear focus on this question in the research project GeodiKon (development of a didactical concept for geometry education).

The first of two mainly addressed fields of geometry research at GeodiKon is the topic “factors”. The second issue is “strategies”: The measurement of spatial intelligence is usually performed with the help of tests. These tests are generally compiled as tests with different groups of tasks and with different types of tasks which address special factors of spatial intelligence. As a result, the different developments of each of the factors of spatial intelligence of an individual can be identified. Those special tasks which address (almost) exactly one factor of spatial ability are called “markers” of the factor [22]. But only if all the probands solve the special marker-tasks with the same intended solving strategies, and the correct factor can be ascertained, valid results of the test can be expected. The recent literature on spatial ability research points out the problem that individuals solve spatial ability test tasks with highly differing strategies (e.g., [47]). This fact opens a wide field of questions, such as: How far can spatial ability tests really address and finally measure the competence of probands at the different factors of spatial ability? and: Which and how many strategies for assessing spatial ability tasks do individuals use?

The focus, therefore, was shifted to the identification and analysis of diverse strategies used for solving spatial tasks. The statements below underpin the intention to put more emphasis on strategies: “The flexible use of strategies or the adequate use of a strategy depending on the given task is a key factor for the optimal solution of spatial tests” (citation translated from [27]), “The amount of strategies and the flexibility in adapting them to the respective task is more relevant for achievement than simple basic cognitive processes” (citation translated from [15, p. 325–326]). Not least, Maier pointed out that “Common alternative solving strategies using other cognitive abilities or different spatial and visual references should thus find recognition” (citation translated from [35, p. 55]). In the research project GeodiKon the topic “strategies” was consciously chosen to answer questions, such as “Which strategies for solving spatial tasks do students use?”, and “Does the training of many different strategies for assessing spatial tasks and thereby the expansion of the individual repertoire of strategies lead to an improvement of spatial ability?”

2. The research hypotheses

The two areas — factors and solving strategies for spatial ability tasks — represent the main focus of investigation within the project GeodiKon. The following two research hypotheses were formulated:

1. Training (making aware, categorising, practicing) of each factor of spatial ability will

lead to an improvement of spatial ability.

2. Training (making aware, categorising, internalizing) of the repertoire of strategies for solving spatial tasks will lead to an improvement of spatial ability.

The huge amount of collected data for the project makes it possible to discuss not only the two research hypotheses, but also a big variety of questions regarding leisure activities and spatial abilities, computer usage and spatial abilities, the accuracy of the ability of pointing to objects, and gender specific effects. Of all these additional results the gender specific effects will be discussed in this paper.

Which of the large number of psychological models for spatial ability should be taken as the scientific basis for this project? During the factorial phase of spatial ability research [39] between 1950 and 1994 many psychometric factor based models of spatial ability were described (e.g., [7, 11, 19, 32, 33, 34, 35, 40, 45, 55]). In a detailed analysis the existing models were compared at the beginning of the project [39]. Maier's approach [35] was formulated as an aggregation of the models existing at that time. Thus, this model served as the starting point of our considerations. Maier took the Thurstone's model [55] with the three factors visualization, spatial relations, and spatial orientation as the basis of his approach. The model of Linn and Petersen [32] with the three factors visualization, spatial perception, and mental rotation turned out to be an outstanding supplement (citation translated from [35]) to the first model. Maier combined these two models and formulated his approach which finally consisted of the five factors visualization, spatial perception, spatial relation, mental rotation, and spatial orientation ([35, p. 51]). Detailed analyses of Maier's approach showed that the four factors visualization, spatial relation, mental rotation, and spatial orientation had also been formulated in three up to nine other models of other researchers. Only the factor spatial perception was just included in the model of Linn and Petersen [32]. The description of this factor according to Linn and Petersen defines the factor spatial perception as the ability to identify the horizontal and the vertical. This very specific ability is considered to be an integrative part of the factor spatial orientation by Thurstone [55]. Thus we do no longer consider the factor spatial perception as a discrete factor. So, Maier's approach — but without the factor spatial percetion — was taken as the scientific basis for the development of the learning material and the test battery in the project GeodiKon.

The factor-based model of spatial ability for the project GeodiKon contains the four factors:

- Visualization
- Spatial Relation
- Mental Rotation
- Spatial Orientation.

Apart from the factors of spatial ability, the second focus of the project GeodiKon lies on different strategies for the solution of spatial ability tasks. It is an aim of the project to provide information about strategies in a structured, focussed, and clear way to students. To be able to realise this goal as a first step we researched existing literature in detail [37]. Many researchers provide lists of spatial ability solution strategies in publications, e.g., Barrat (key features strategies, move object strategies, and move self strategies) [2], Just and Carpenter (mental rotation around the global coordinate system, mental rotation around a user coordinate system, compare the characteristics of objects with one another, and change of perspective) [26], Dünser (moving oneself or moving the object, concentration on details or on the whole, and reflection and visualization) [9, p. 159], and Schultz (mental rotation strategy, perspective change strategy, analytic strategy) [46]. The analysis of strategy research showed that four antagonistic pairs of strategies for assessing spatial tasks can be identified. These eight strategies do not imply any claim of completeness, although many researchers (see above) recognize these eight strategies or subsets of them as THE relevant strategies [37]. In addition to the four pairs of strategies described below, there are further terms frequently formulated: avoidance strategies, complementary strategies, mixed strategies, verbal-analytical strategies, and logical consequential thinking [17, 35, 47], which after close analysis can be regarded as parts of one of the pairs of strategies. The structured model of the "four pairs of strategies for the solution of spatial ability tasks", which was used at the project GeodiKon, is as follows:

1. Holistic Strategy — Analytic Strategy
2. Spatial Thinking — Planar Thinking
3. Move Object — Move Self
4. Verifying Strategy — Falsifying Strategy

The individual pairs of spatial ability solution strategies form antagonistic pairs. In tests geometrical objects are generally comprehended either holistically or analytically. Test persons either construct a mental spatial model of the objects depicted (spatial strategy) or they start from the plane image of the object (planar thinking). When solving spatial ability tasks probands are positioning themselves outside the scene or, conversely, test persons — particularly in tasks of spatial orientation — put themselves into the proposed setting and mentally move around the objects. Test persons in general prefer verifying or falsifying procedures in solving the given tasks. If there are several acceptable solutions, they either try to find the right solution straight away or they exclude false solutions one by one until only one solution is left as the correct one. A detailed argumentation of this model, characteristics of the strategies, and cross connections between the strategies is discussed at [37].

3. Description of the study

The research project GeodiKon, funded by the Federal Ministry for Education, the Arts and Culture (BMUKK-20.040/0012-I/7/2012) and the University of Education of Salzburg, was carried out from 2013 to 2014 in the three Austrian provinces of Salzburg, Styria, and Lower Austria. The project partners were the private University of Education of Lower Austria and Vienna, the University of Education of Styria, the University of Innsbruck, the University of Salzburg, the University of Vienna, the Technical University of Vienna, and the Workgroup for Didactical Innovations (ADI). The major aim of the project was to provide results regarding the two research hypotheses. Therefore the following milestones were defined:

1. Development of specific (paper and pencil) learning material for 12 weeks of training of the four factors visualization, spatial relations, mental rotation, and spatial orientation of spatial ability [32, 35, 37, 55] with the aim to train a balanced and extensive development of the learners' spatial ability.
2. Development of a structured model of strategies for solving spatial tasks [2, 26, 46] with the aim to extend the learners' repertoire of problem solving strategies.
3. Development of the battery of tests, execution of the pretests and posttests, in-service training of the teachers of the project classes how to use the learning material, and the information about strategies, and analyses of the project data.

4. Providing a user-friendly book with all the specific learning material and the results of the project for lecturers for mathematics and geometry.
5. Training for teachers and lecturers how to use the material in classes.
6. Dissemination of the results of the project in conference presentations and papers.

The project was carried out in a pretest-posttest-design. During the first phase of the project (January until September 2013) the project team compiled learning material for 12 weeks of lessons in geometry and mathematics. The learning material contains specific spatial ability tasks, training the factors of spatial ability and the different strategies for solving spatial tasks. The structured model of the four pairs of strategies (holistic strategy, analytical strategy, spatial thinking, planar thinking, move object, move self, verifying strategy and falsifying strategy) for the solution of spatial tasks was developed and the test battery was set up. In September and October 2013 the pretests were carried out. Directly after the pretests the twelve-week long learning phase began for the treatment groups. In January and February 2014 the posttests were carried out in all school classes. March until November 2014 was the time for digitalization, preparation and analyses of the collected data, compilation of the user-friendly book with all the special learning material [38], training for teachers and lecturers how to use the material in classes, and dissemination of the results of the project in conference presentations and papers.

Altogether 46 test classes with 903 students at the age between 12 and 14 years from all types of secondary school (Hauptschule (HS), Neue Mittelschule (NMS), Bundesrealgymnasium (BRG), and Bundesgymnasium (BG)) in the three Austrian provinces of Salzburg, Styria, and Lower Austria took part in the project. The invitation to take part in the project was sent as a digital newsletter to 2.260 teachers at BG/BRG and HS/NMS (606 at BG/BRG and 1.654 at HS/NMS). This newsletter periodically addresses teachers of geometry in the German speaking area (mainly Austria). Originally the project was designed for 10 classes. Because of the great interest (96 teachers and their classes) we finally took 46 classes to take part in the project. The selection criteria were: Residents of one of the three provinces Lower Austria, Styria, or Salzburg, and a balance of distribution of all probands regarding their sex, age, school type, and rural and city schools. The 12 project classes of Salzburg, the 12 project classes of Styria, and the 22 project classes of Lower Austria had been supervised by province coordinators (two persons in Styria, one person in Salzburg, and two persons in Lower Austria). Those five persons were the test supervisors for all the pretests and posttests, and worked with exactly the same time schedule for the tests. All the teachers of the project classes also had training sessions in which they learnt how to work with the learning material and how to provide the information about the different strategies to solve spatial tasks to the students. The sessions were organised to make sure that all the classes would work in (nearly) the same way during the 12 weeks of the treatment. For half of a lesson (about 25 minutes) the project classes worked with the special learning material during the treatment period (12 weeks) in the subject "Geometrisches Zeichnen". During the second half of a lesson the teachers worked with their classes independently of the project. Directly before the treatment period all the students took the pretests and directly after the treatment they took the posttests. After the posttests all the data were collected and especially the difference in the performance of the students was analysed. According to the classification of spatial training studies from Newcombe et al. [42] GeodiKon was set as a general training study (trained a variety of spatial tasks for different spatial tests) and a long duration study because it lasted for at least a whole semester.

4. The test battery

The test battery of the pretests and the posttests consisted mainly of four spatial ability tests Three-Dimensional Cube Test (3DW) [14], Differential Aptitude Test (DAT) [3], Mental Rotation Test (MRT) [43] and Spatial Orientation Test (SOT) [23]. In addition to the four tests their were different questions about which strategies students used to solve spatial tasks and questions about age, gender, computer usage, leisure activities, school marks in Mathematics, German and English, and type of learner. The pretests lasted for 85 minutes and the posttests for 77 minutes. We wanted to know with which strategies probands solved the tasks of the four spatial ability tests. So after each of the four tests the students once again got one of the tasks, which was arbitrarily chosen. When the students solved the task, they were asked to observe themselves accurately with which spatial strategy they solved the task. Then students answered the four questions concerning the different strategies they used from the model of the four pairs of strategies — each in an eight-part scale (Table 1).

Table 1: The four questions for students concerning the four pairs of solving strategies for spatial tasks

1) Looked at the object in its entirety: You looked at the whole object. You did not concentrate on parts of the object only. You visualised the whole object and found the solution right away.	Looked at parts of the object: You concentrated on parts of the object only. You did not have to use the whole object for the solving process.
o o o o	o o o o
2) Spatial thinking: You created a mental, three-dimensional model of the object and solved the task by working on this mental model.	Planar thinking: You created a mental, planar (two-dimensional) model of the setting and solved the task by working on this planar mental model.
o o o o	o o o o
3) Move self: You placed yourself inside the setting and moved around mentally and changed your perspective.	Move object: You positioned yourself mentally as an observer outside the setting and moved the objects.
o o o o	o o o o
4) Falsifying strategy: You identified all the incorrect solutions first and excluded them step by step.	Verifying strategy: You had the correct solution in mind and worked on it directly.
o o o o	o o o o

For better understanding and traceability of the pretest results of the project, the four spatial ability tests which we used are explained as follows. Each of the test addresses specific factors of spatial ability. The Three-Dimensional Cube Test (3DW) addresses the factor visualization, the Differential Aptitude Test (DAT) the factors visualization and spatial relations. The Mental Rotation Test (MRT) focuses on the factor mental rotation and, finally, the Spatial Orientation Test (SOT) addresses the factor spatial orientation. These assignments had been specified in the best possible way. They do not raise the claim to be fully selective and accurate. These characteristics were not intended because the analysis will not go into detail regarding the varying improvement of the four factors.

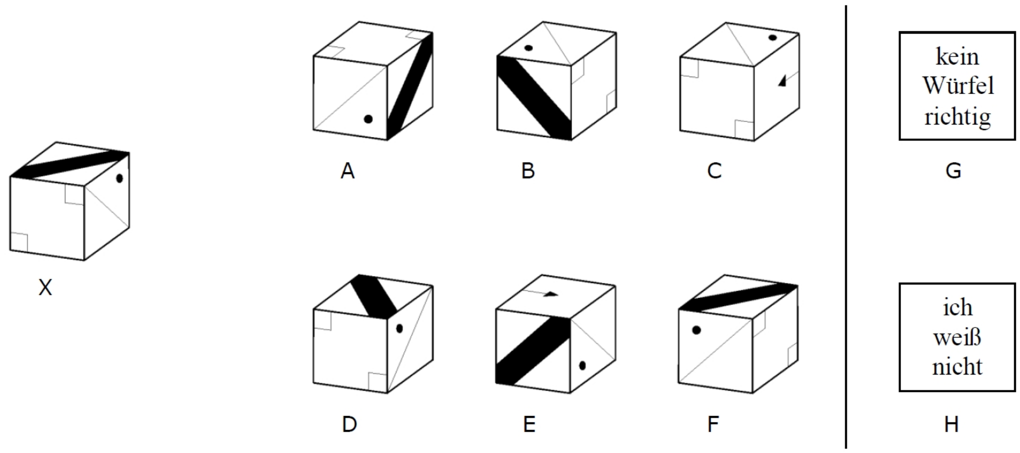

Figure 1: An example of the Three-Dimensional Cube Test (3DW)

4.1. Three-Dimensional Cube Test (3DW) [14]

This test investigates whether the cube A, B, C, D, E or F is exactly the same cube as the cube X or whether the right answer is G (no cube matches). If you do not know the solution, you have to choose H (I do not know the answer). Each pattern at the side faces of the cube occurs only once. The test author G. Gittler provided a special version of the 3DW-test for this project with 13 tasks. The first one is a hidden warming up task and is not being counted. The test lasts for 15 minutes.

4.2. Differential Aptitude Test (DAT) [3]

The tasks of this test consist of handling folding templates with shades and patterns. The templates can be folded to three dimensional objects. Each task shows one folding template and four three dimensional objects. You have to choose which of these three dimensional objects A, B, C or D can be folded from the folding template provided. The test consists of 15 tasks and lasts for 8 minutes. At each task exactly one answer is correct.

4.3. Mental Rotation Test (MRT) [43]

An object is presented on the left. The probands have to determine which two of the four sample stimuli A, B, C, and D on the right are rotated versions of the target stimulus [43].

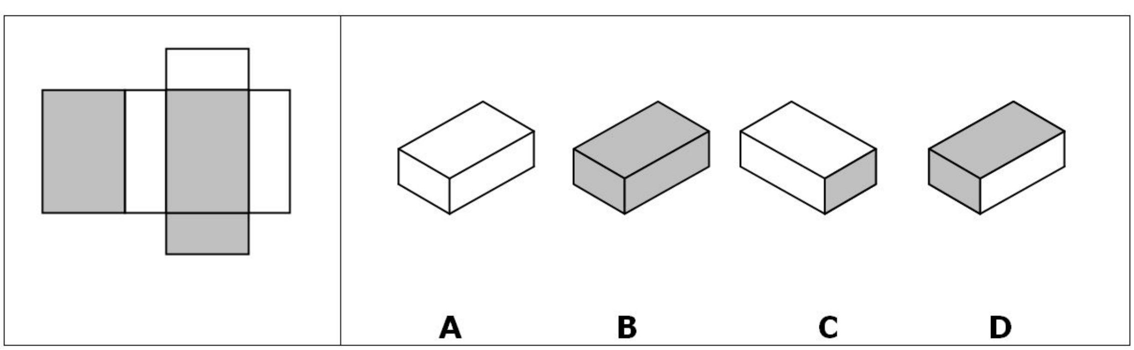

Figure 2: An example of the Differential Aptitude Test (DAT)

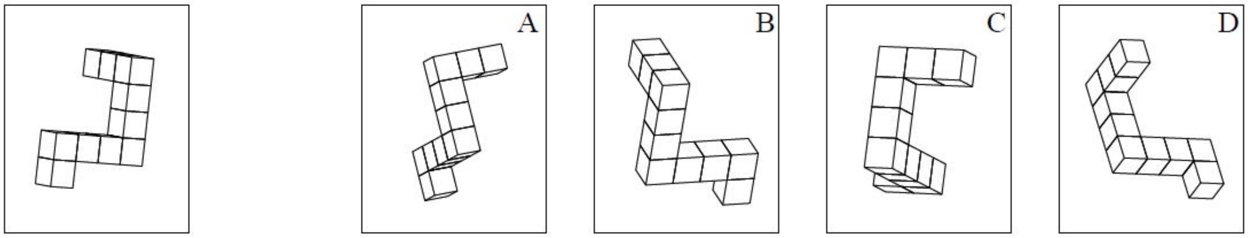

Figure 3: An example of the Mental Rotation Test (MRT)

A task is solved correctly if exactly both correct answers are marked. Only then the proband gets one point. The tests consists of 24 tasks and lasts for 6 minutes.

4.4. Spatial Orientation Test (SOT) [23]

This is a test of one's ability to imagine different perspectives or orientations in space. In each task one can see a picture of an array of objects and an "arrow circle" with a question about the direction between some of the objects. For each question one should imagine to be standing next to one object in the array (which is named in the center of the circle) and facing another object, named at the top of the circle. The task is to draw an arrow from the center object showing the direction to a third object from this facing orientation [29]. In this test you can get no points. At each task the deviation angle from the right answer is measured. The angle is measured not orientated, so therefore all the deviation angles are in the range between 0° and 180°. The SOT consists of 12 tasks and lasts for 8 minutes.

5. Results

903 students from the 46 project classes participated in the tests. Of those 903 students 786 students were present at both tests. There were no systematic errors, so we accepted missing completely at random (MCAR). Of the 786 students who took part in both tests 771 students delivered evaluable tests. Those 771 students (413 male and 358 female) formed the basis of

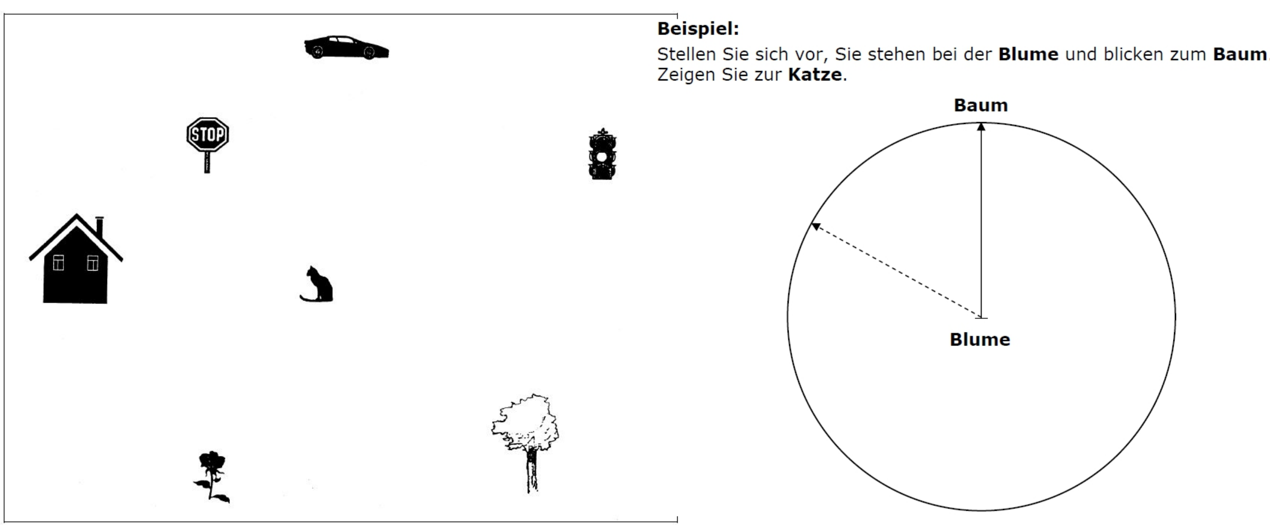

Figure 4: An example of the Spatial Orientation Test (SOT)

the data analyses. The analyses of the 771 data sets showed that by examining the internal consistency of the used tests in the overall sample all the reliabilities were above the required value of 0,7 [28]. So all the tests used can be seen as sufficiently consistent.

All the probands were at the age between 12 and 14 years. There are two reasons why this is quite an interesting age. On the one hand because according to Beanninger and Newcombe [1] there exist nearly no spatial ability studies with children at this age or younger — especially spatial ability studies like GeodiKon where the activity preferences of children is focussed and the correlations to their spatial ability is examined. And on the other hand it is exactly the age where Voyer et al. [57] report first sex differences on the DAT. According to Voyer et al., the earliest age at which sex differences were reported at different spatial tests is between 7 and 14 years. So, GeodiKon studied sex differences within children quite on the edge of the age-interval were it first can be recognised. The 46 project classes were divided into four groups, first of all to have the possibility to analyse the global differences between the three treatment groups A, B, and C and the control group D, and secondly to find out if different developments of the subgroups A, B, and C can be observed. The test analysis (dichotomous logistical model of Rasch) (1-PL) were made with the three tests 3DW, DAT and MRT. All the detailed analysis carried out by E. Svcenik can be downloaded at [51]. They were executed with the software R eRm from Mair et al. [36].

The Andersens Likelihood-Ratio-Test with a χ^2 of just 14.35 from 11 degrees of freedom (p = 0.214) shows no indications of a deviation of the 3DW-test from the dichotomous logistical model of Rasch. So therefore we can take the 3DW-test as Rasch conform. The ICC plot in Figure 5 shows this clearly in a visual way. The other two tests (DAT and MRT) are not Rasch conform although the ICC are nearly parallel (see all the plots at [51]). The posteriori contrasts and Student t-tests can be downloaded at [52]. Further statistic descriptions and

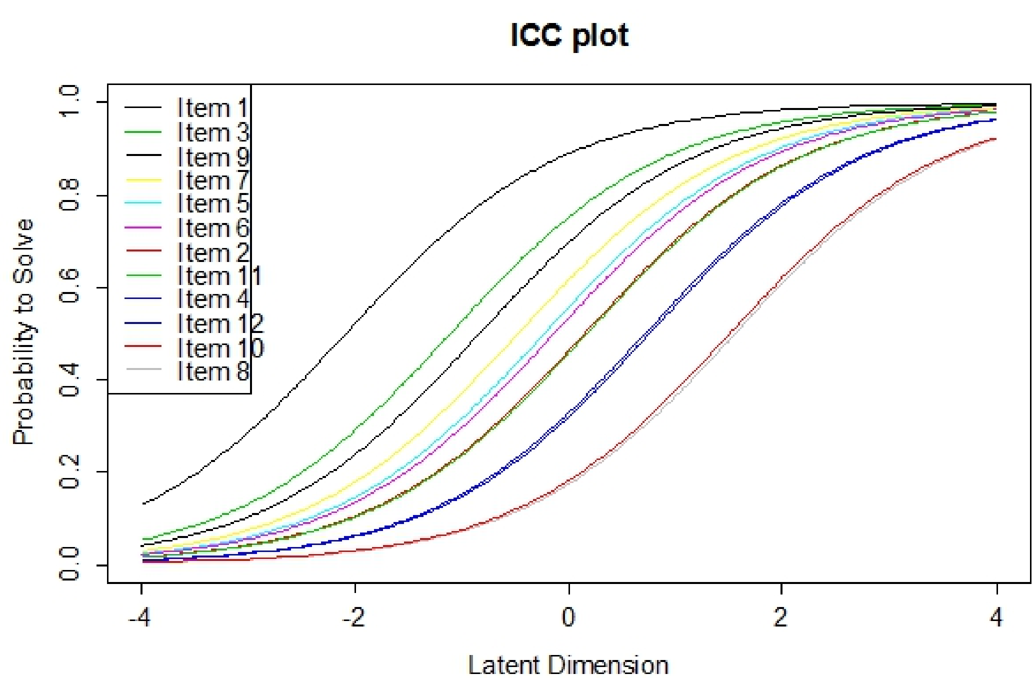

Figure 5: ICC-Plot (Item Charateristic Curve-Plot) of the Rasch analysis from the 3DW-test. The test had 13 items. The first one was a hidden warming up item. So just the 12 actual test items were examined and plotted.

exhaustive analyses can be downloaded at [49] and [50].

The four groups (treatment groups: A, B and C; control group: D) of the project:

- Group A (163 students with valid data records; treatment group): The students of the group-A-classes got information about strategies for the solution of spatial tasks for 12 weeks (intervention period) and they worked with the newly developed learning material for 25 to 35 minutes per week
- Group B (267 students with valid data records; treatment group): The students of the group-B-classes worked for 25 up to 35 minutes per intervention week with the newly developed learning material (but got no information about strategies for the solution of spatial tasks)
- Group C (189 students with valid data records; treatment group): The students of the group-C-classes were students who were taught geometry as a separate subject in their school curriculum for at least between one to two lessons per week. They did not work with the developed learning material and got no information about strategies for the solution of spatial tasks
- Group D (152 students with valid data records; control group): The students of the group-D-classes were students of the control group. They did not work with the developed learning material and got no information about strategies for the solution of spatial tasks and they even had no separate subject "Geometry" in their school curriculum)

The students of the treatment groups A and B worked with the learning material for 25 to 35 minutes each week of the 12 long week long interventions phase. The learning material consist of more than 100 geometric tasks, which are training especially the four different

Table 2: Arithmetic Mean (AM) and Standard Deviation (SD) of the results from the pretests, posttests, and the difference between pretests and posttests

	A	B	C	D
	$162 \leq n \leq 170$	$263 \leq n \leq 270$	$187 \leq n \leq 191$	$150 \leq n \leq 155$
		3DW-test		
Pre AM/SD	3.47/2.71	4.22/2.65	3.47/2.64	3.44/2.61
Post AM/SD	4.61/2.83	5.46/2.89	4.63/2.87	4.25/2.95
Diff AM/SD	1.14/2.27	1.24/2.48	1.16/2.45	0.81/2.35
		DAT		
Pre AM/SD	7.57/3.54	8.79/3.24	7.94/3.32	8.33/3.36
Post AM/SD	8.71/3.75	10.35/3.71	9.83/3.57	9.45/3.71
Diff AM/SD	1.14/3.06	1.56/2.97	1.90/2.75	1.12/2.95
		MRT		
Pre AM/SD	8.71/4.52	9.12/4.60	8.39/4.17	8.19/4.57
Post AM/SD	12.07/5.69	13.80/5.61	12.20/5.33	11.31/5.28
Diff AM/SD	3.36/4.64	4.68/4.31	3.81/4.34	3.12/4.48
		SOT		
Pre AM/SD	65.98/30.32	57.63/29.94	64.86/30.37	59.43/26.95
Post AM/SD	56.56/31.11	48.50/29.57	57.12/30.89	52.62/28.92
Diff AM/SD	-9.42/24.37	-9.12/22.95	-7.74/21.68	-6.81/21.02

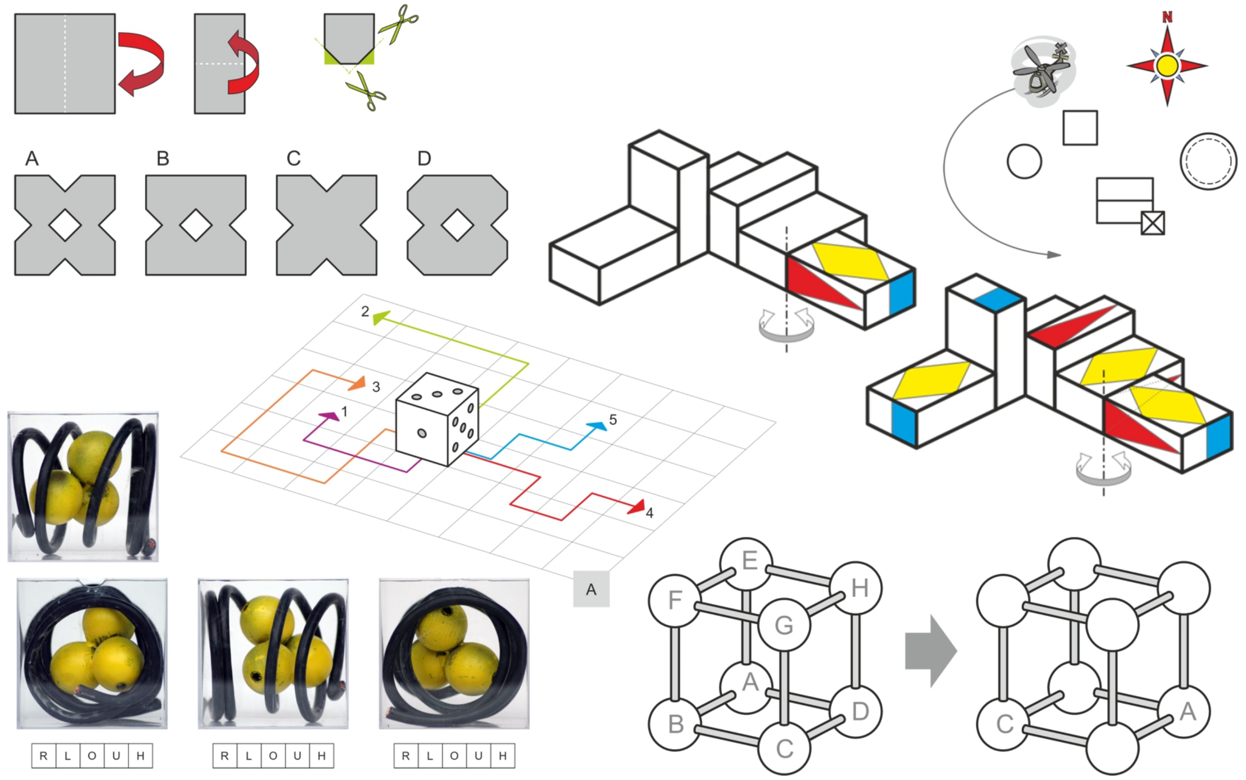

Figure 6: Some few examples of the learning material from GeodiKon

factors of spatial ability in a playful way for students between 10 and 19 years. All the tasks should be solved by the students without the help of the teacher. The whole school lesson lasts for 50 minutes in Austria. So about half of the lesson in the subject "Geometrisches Zeichnen" was used to work with the special material for 12 weeks. During the second half of the lesson the teachers worked with their classes independently of the project. The most of the tasks of the learning material can be solved just by using a pencil. Some of the tasks need coloured pencils. Students generally do not need a ruler or any other equipment (e.g., calculator) to edit the tasks. Figure 6 shows some examples of the tasks of the learning material. After the intervention phase all the tasks of the project were compiled to a user-friendly book [38] with additional recommendations for teachers how to use the material and a table which provides the information of which task trains which factor.

All the teachers of the project group A (strategies and learning material) and group B (learning material) had special training sessions in which they learnt how to work with the learning material and how to provide the information about the different strategies to solve spatial tasks to the students. The sessions were organised to make sure that all the classes would work in (nearly) the same way during the 12 weeks of the treatment. Regarding to the strategies the teachers got many information about the scientific background of the model of the four pairs of strategies to solve spatial tasks. They were also trained to provide lessons to students where they explain the eight strategies in many different ways (e.g., how to solve the same tasks in different ways with the help of different strategies). The teachers got no explicit storyboard how to introduce strategies in the same way. They just should make sure that students of the classes from group A were thinking about different strategies in every lesson and so therefore hear about new strategies, hear about new details about well known strategies and finally widen the repertoire of strategies. In addition to all that students got a compact poster where the model of the four pairs of strategies to solve spatial tasks was introduced and each strategy was explained in an student friendly way.

Table 2 shows the performance of the students in all four project groups (treatment groups: A, B, and C; control group: D) in the pretests and the posttests by means of the arithmetic means (AM) and and standard deviation (SD). It also presents the descriptive statistical differences of the results from the pretests and the posttests, which corresponds with the improvement of the students. Figure 7 shows the data of Table 2 graphically. When we look at the diagrams, we can assume that the three treatment groups A, B, and C had a very similar improvement of their performances. We also see that the control group D has a lower improvement in each of the four tests as the three treatment groups. The statistical proof of these assumptions shows that the simultaneous comparison of the treatment groups and the control group renders no effects (3DW-test: $F_{3;36,36}$= 0,661; p = 0,582; DAT: $F_{3;44,04}$=1,40; p = 0,257; MRT: $F_{3;42,09}$= 1,49; p = 0,231; SOT: $F_{3;41,39}$= 0,46; p = 0,715) [50]. It must be stated that there is no simple and direct answer to the research hypotheses (see chapter 2). Many results from the detailed analyses provide a lot of findings which will be discussed below.

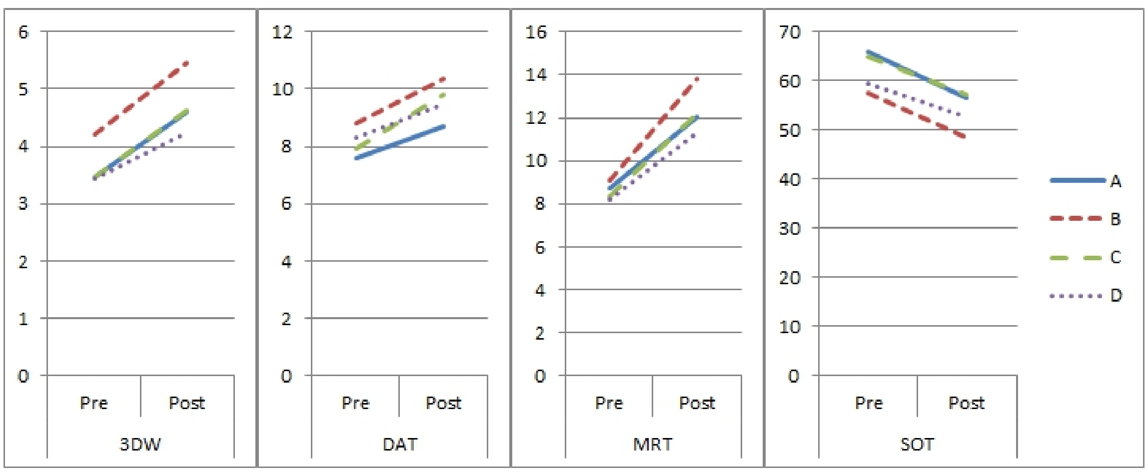

Figure 7: Improvement of all four project groups A, B, C, and D

5.1. Geometry education and geometry lessons

The three treatment groups A, B, and C were combined together to one group because they showed very similar improvement in performance (Figure 8).

All the students of this new group attend the subject "Geometrisches Zeichnen" (GZ). Thus, the new GZ-group enables the analyses whether the subject GZ has an impact on the increase of spatial ability. Figure 8 shows the development of the both groups (with GZ and without GZ; group "without GZ" = control group D). The arithmetic means (AM) which are the basis of Figure 8 are displayed in Table 3.

With inferential statistical methods two questions regarding the impact of GZ lessons on spatial ability were discussed:

- Effect "GZ": Do students who have GZ lessons perform better in the four spatial ability tests than others who do not have GZ lessons, independent from the time of measurement?

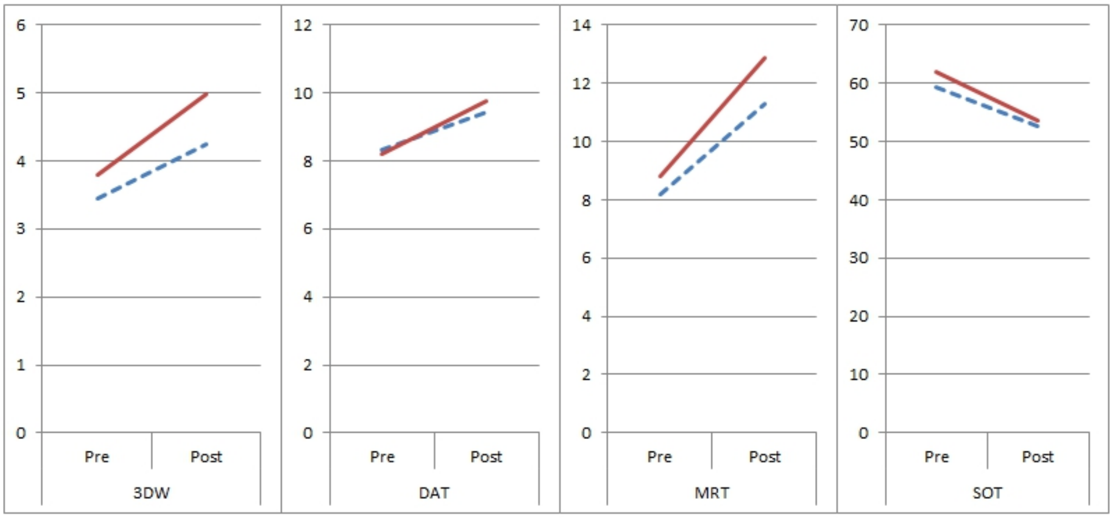

Figure 8: Improvement in performance of the two groups "without GZ", and "with GZ"

Table 3: Arithmetic means (AM) of the improvement in performance of the two groups "without GZ", and "with GZ"

	Without the subject GZ	*With the subject GZ*
3DW Pret/Post	3.44/4.25 items solved	3.79/4.99 items solved
DAT Pret/Post	8.33/9.45 items solved	8.20/9.75 items solved
MRT Pre/Post	8.19/11.31 items solved	8.79/12.85 items solved
SOT Pre/Post	59.4°/52.6° deviation angle	62.0°/53.6° deviation angle

- Effect "Potential GZ": Do students with GZ lessons have a higher improvement of their performance than students from the control group (= group without GZ)?

Table 4 shows the two effects "GZ" and "Potential GZ" in the four spatial ability tests 3DW-test, DAT, MRT, and SOT. The last line of Table 4 presents the general increase in performance of all students (without differentiation in subgroups). The text in the cells of statistic significant effects is shown in bold letters.

In all four tests a highly significant increase of performance during the intervention phase is recognizable. Many effects can be made responsible for this unspecific trend: learning

Table 4: The increase in performance in the four tests

Effect	3DW-test	DAT	MRT	SOT
GZ	$F_{1;764}$=**5,502** **p=0,019**	$F_{1;784}$=0,096 p=0,757	$F_{1;784}$=**6,929** **p=0,009**	$F_{1;771}$=0,419 p=0,518
Potential of GZ	$F_{1;764}$=2,944 p=0,087	$F_{1;784}$=2,683 p=0,102	$F_{1;784}$=**5,606** **p=0,018**	$F_{1;771}$=0,941 p=0,332
Increase in performance	$F_{1;764}$=**83,901** **p<0,001**	$F_{1;784}$=**101,968** **p<0,001**	$F_{1;784}$=**323,937** **p<0,001**	$F_{1;771}$=**59,191** **p<0,001**

effects, training effects because of the test repetition, ripening process effects, development process effects, treatment effects, and combinations of these effects [13, 51]. Students who have GZ lessons show significantly higher performances already at the start of the treatment in the 3DW-test ($F_{1;764} = 5{,}502$; p = 0,019) and in the MRT ($F_{1;784} = 6{,}929$; p = 0,009). There is no such effect recognizable in the DAT ($F_{1;784} = 0{,}096$; p = 0,757) and in the SOT ($F_{1;771} = 0{,}419$; p = 0,518). The Potential GZ Effect provides information if students with GZ lessons have a higher improvement of their performance during the 12 week long intervention phase than students from the control group (= group without GZ). In all four spatial tests higher performances of the GZ group than in the control group can be recognized. The increase of performance is significant in the MRT ($F_{1;784} = 5{,}606$; p = 0,018).

5.2. Teacher effect, class effect

For the analyses of the project the probands were not treated as individuals, they were grouped in their classes (cluster sampling). This way of sample drawing does not allow to estimate the standard error because in general it can be expected that students from one class do have a more homogenous increase of performance than individuals from the whole sample. The multi-level structure of the data is considered at the analyses with a Linear Mixed Model Approach [51, 53, 58]. To decrease the amount of parameters which have to be estimated, the increase of performance was defined as a dependent variable. For each of the four tests the Inter Class Correlation (ICC) was calculated [50]. The analyses point out that in the 3DW-test (ICC von 5,20%; Wald-Z = 2,065; p = 0,039), in the DAT (ICC von 5,36%; Wald-Z = 2,387; p = 0,017), and in the MRT (ICC von 14,29%; Wald-Z = 3,46; p = 0,001) a substantial class effect is recognizable. So in these three tests students from one class do perform more homogenously [50]. The highest effect can be noticed in the MRT. In contrast to these results there is no class effect recognizable in the SOT. The SOT measures individual performances where the usual effects (e.g. class effect, teacher effect, classroom management) do not matter [50].

5.3. Three-dimensional cube test (3DW-Test)

The analysis of the 3DW-test focussed on so-called ability parameters from the dichotomous logistical model of Rasch (1980). Therefore some steps were needed to prepare the data.

- The first task of the 3DW-test is a hidden warming up task and is not taken into consideration in the analysis.
- Only those tests were analysed in which students edited at least eight items (incl. the hidden warming up item), because only then we had enough reliable information.
- Depending on the number of correctly solved items (= test raw value RW) a so-called raw value estimator (RWS) was calculated. This RWS estimates in a best possible way how many items a student would have solved if she/he had edited all 12 items of the 3DW-test.
- The RWS are the test scores of each proband, because only then the results can be compared in a realistic way even if the students did not edit the same amount of items.
- The last step is to convert the RWS into so called ability parameters (PAR) according to the Rasch model. Now we have an interval scale level and because of this we can use statistic tests for the further analyses [13].

Gittler provided in [13] a conversion table with all the ability parameters (PAR). In the 3DW-test (and also in the other three tests) a highly significant increase of performance is identifiable (see Table 4). The differentiation between the four project groups A, B, C, and D shows that this is also valid for each of the four groups. The increase of performance is shown in Table 5 as the difference of the ability parameters between pretest and posttest. The effects can be interpreted as small to medium size according to Cohen's d (Table 5), in which in general an effect of d = 0.20 is a small effect, an effect of d = 0.50 is a medium size effect, and d = 0.80 is a large effect [13].

Table 5: Increase in performance of all groups at the 3DW-test

Group	*Ability parameter (PAR)*	*Cohen's d*
A / B / C	0.61 / 0.60 / 0.63	0.418 / 0.436 / 0.437
D	0.41	0.270

It should be noted that the increases in performance of the treatment groups A, B, and C are quite similar and that the increase at the control group D is much lower (see Figure 9).

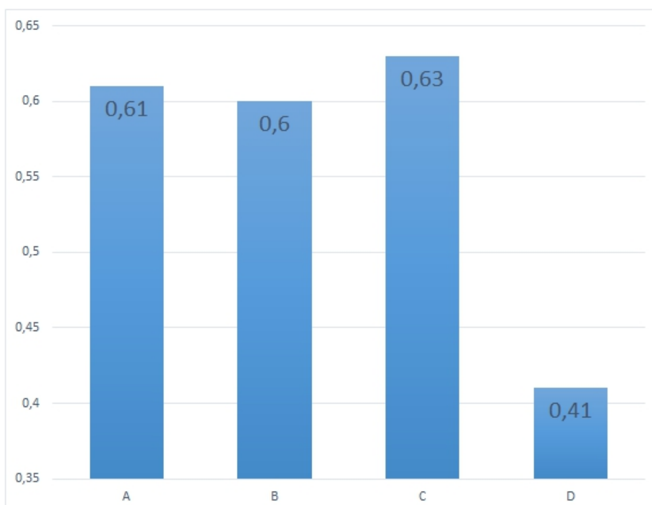

Figure 9: Increase in performance of the groups A, B, C, and D at the 3DW-test measured in ability parameters (PAR) from Table 4

Below the increase of performance at the 3DW-test of all groups is presented separately according to gender. In all groups we can again recognize a significant increase of performance. Table 6 shows the improvement of all groups in the 3DW-test (m = male; f = female; N = number of students of each group; PAR = ability parameter).

In all three treatment groups A, B, and C the girls perform better than the boys. It is remarkable that this is exactly the other way round in the control group D, in which the male students have a greater increase of performance than the female students (see Figure 10).

This interesting observation indicates that general analyses of all data regarding to differentiation by gender could deliver valuable results.

Table 6: Gender-specific improvement of all groups in the 3DW-test

Group	*Sex*	*N*	*PAR*	*Cohen's d*
A	m / f	97 / 58	0.60 / 0.66	0.403 / 0.468
B	m / f	160 / 104	0.59 / 0.63	0.431 / 0.456
C	m / f	89 / 96	0.52 / 0.72	0.376 / 0.473
D	m / f	57 / 92	0.44 / 0.36	0.316 / 0.222

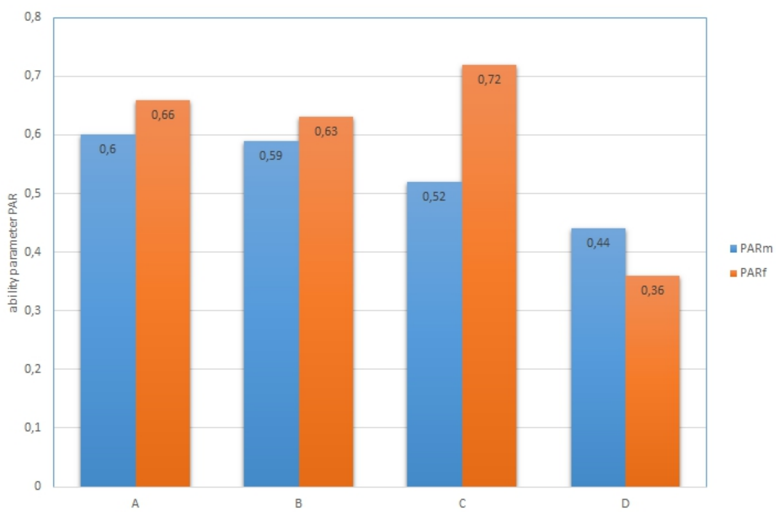

Figure 10: Increase of performance of the groups A, B, C, and D in the 3DW-test shown in ability parameters (PAR) and differentiated by gender (PARm = male (left and blue); PARf = female (right and yellow)

5.4. Gender sensitivity

The gender sensitivity analyses were carried out for each test and are listed test by test below. In the 3DW-test the analysis shows that the t-test for male probands (t = 0.762, df = 401 and p = 0.447) is not significant. Also the Cohen's d effect is very small (0.11) for male students. For female students the t-test is significant t = 1.972, df = 348 und p = 0.049. The Cohen's d effect shows a small effect (0.24). The gender sensitivity analysis in the 3DW-test shows that there is a significant treatment effect (p = 0.049) for female students, which can be seen as a small Cohen's d effect [13]. 11.8% (p <0.001) of the variance of the DAT can be explained with the variables gender, school type, and school level, whereby gender does not show any effect [49]. The effects of the MRT (16.4%; p <0.001) can be explained with the variables gender, school type, and school level. In this test male students score 2.55 points more than female students. Male probands have an average of 9.82 correct answers, which is clearly more (F = 45.68; p<0.001) than the female students, who have 7.27 correct solved items. Male students (11.31) also edited more items than female students (10.90) — no matter if the item was solved correctly or not (F = 9.24; p = 0.002) [49]. In the SOT the average of the deviation angle of the correct solution was 59.04° in the pretests and 50.64° in the posttests. So the average reduction of the deviation angle from the correct answer was 8.40°, which means a highly significant value for all classes and groups ($F_{1;44.99}$ = 80.56; p <0.001). The deviation angle of male students is smaller than the deviation angle of female students in the pretests

(14.4°) and in the posttests (13.5°). This shows that the gender difference stays nearly the same in the pretest and posttest and no significant gender specific treatment effect in the SOT ($F_{1;757}$= 0.28; p = 0.597) can be seen. The analyses are in accordance with scientific literature. Gender specific differences regarding spatial ability seem to be much smaller than was supposed in former days. They nearly cannot be detected in the factors visualization and spatial relations. They only exist in the factor mental rotation (only in speeded tests) and in the factor spatial orientation [16].

Female and male students used different strategies in the four spatial ability tests:
3DW-test: Gender-specific differences in the 3DW-test can only be stated for the falsifying strategy. This strategy is used much more often by female students than by male students (F = 6,92; p = 0,009).
DAT: In the DAT we can see the same effect as in the 3DW-test. The falsifying strategy is used much more often by female students than by male students (F = 17,64; p<0,001).
MRT: In the MRT two effects are significant. Female probands much more use the falsifying strategy than male probands do (F = 8.72; p = 0.003). Male students do use the holistic strategy much more often than female students (F = 15.85; p <0,001).
SOT: The use of the falsifying strategy is hardly recognisable in this test. So this strategy is rarely used in this test. However, this strategy is used much more often by female students than male students (F = 4.43; p = 0.036) (cf. [49]).

5.5. Which strategies for solving spatial tasks do students use?

All the questions regarding the strategies used in the four tests were answered in differing ways by the probands, which is an indication that the students dealt with this issue seriously. In all the four tests different strategies were used in different variations and combinations. Especially in the SOT the strategy move self was used much more often than in the other three tests, in which the probands more often moved the object mentally. As expected, in the SOT the students visualised the scene as a plane scene and therefore used the strategy planar thinking much more often than in the other three tests. In the MRT the probands much more often used the holistic strategy than in the other three tests. The students used very similar strategies in the DAT and in the 3DW-test (cf. [49]). Figure 11 provides an overview of the strategies used (ordered by the model of the four pairs of strategies for assessing spatial ability tasks) by the probands in the four spatial ability tests 3DW-test, DAT, MRT, and SOT. Each bar is divided into eight rectangles. Each rectangle specifies the percentage of how many probands marked the first value of the eight-parted scale, the second value, the third value and so on.

5.6. Change of strategies from the pretests to the posttests

The focus of these analyses was put on finding out how the students changed their solving strategies from the pretests to the posttests. Therefore the Inter Class Correlation (ICC), which explains which part of the variance can be explained as a class effect, was calculated. It was assumed that there is a connection between the strategies used and the class teacher. So we expected that students of a class would change their strategies in a similar way. Table 7 shows that this assumption cannot be confirmed. For all strategies the Inter Class Correlation is in a low one-digit percent range and therefore (with the exception of one value in the 3DW-test) far away from significance. Therefore we can note that the change of strategies is an individual characteristic and there are no class effects noticeable [50].

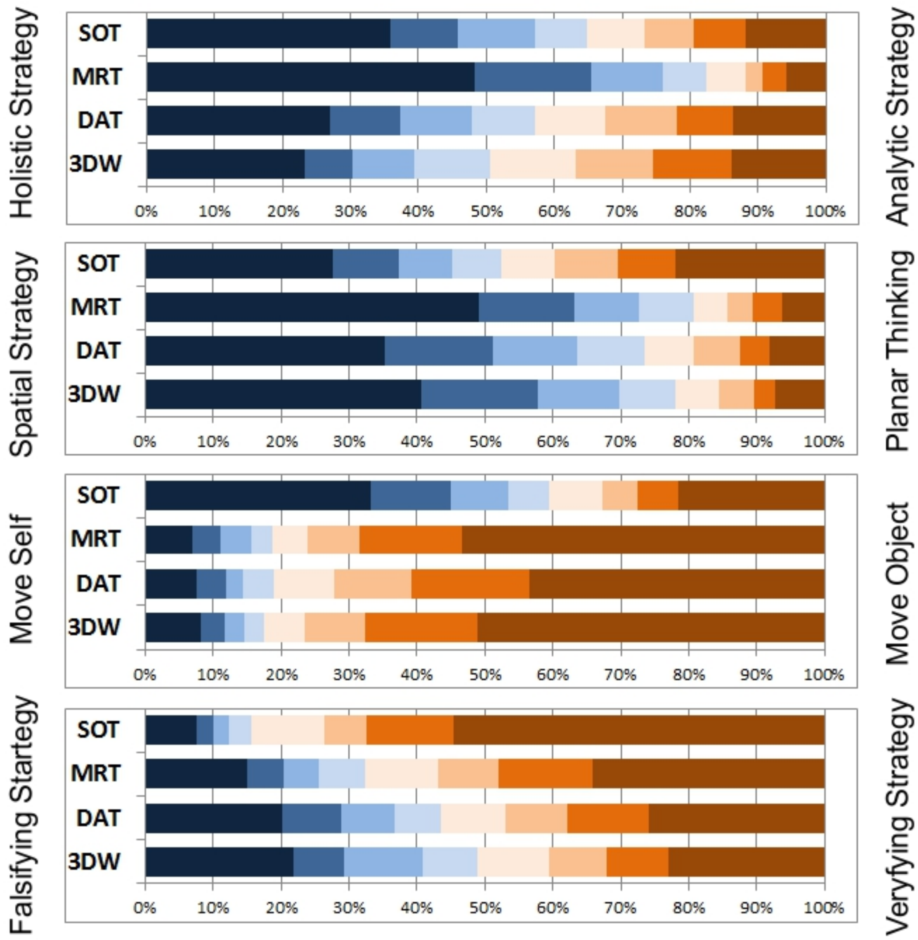

Figure 11: Overview of the strategies used by the students in the four spatial ability tests 3DW-test, DAT, MRT, and SOT

Therefore the effects of the treatment groups and the control group can be analysed by multivariate analysis of variance.
3DW-test: A highly significant change of the solving strategies used in the 3DW-test can be determined (Pillai-track = 0.065; $F_{4;694} = 12.026$; p <0.001). In the posttests the students much more used the holistic strategy and the move object strategy.
DAT: Also in this test a highly significant change of the solving strategies used can be determined (Pillai-track = 0.073; $F_{4;682} = 13.491$; p <0.001). We can see that the probands much more used the holistic strategy and the move object strategy in the posttests.
MRT: As in both tests above, also here we can determine a highly significant change of the solving strategies used in the MRT (Pillai-track = 0.061; $F_{4;706} = 11.497$; p <0.001). Here the students changed from the move self strategy in the pretests to the move object strategy in the posttests.
SOT: Finally, also in the SOT a highly significant change of the solving strategies used can be noted (Pillai-track = 0.020; $F_{4;673} = 3.518$; p = 0.007). Probands more often used the analytic strategy and the planar strategy in the posttests [50].

Only the students from treatment group A got information about different strategies for solving spatial ability tasks. Can we observe a different behaviour for the changes of strategies

Table 7: Inter Class Correlation (ICC) of the strategy changes of students for all four tests

Pair of Strategies	*ICC*	*Wald-Z*	*p*
3DW-test			
Holistic – Analytic	2.31%	1.176	0.240
Spatial – Planar	0.53%	0.390	0.696
Move Self – Move Object	2.22%	1.272	0.203
Falsifying – Verifying	4.48%	1.981	0.048
DAT			
Holistic – Analytic	0.41%	0.281	0.779
Spatial – Planar	2.66%	1.544	0.123
Move Self – Move Object	0.96%	0.627	0.531
Falsifying – Verifying	2.84%	1.473	0.141
MRT			
Holistic – Analytic	0.41%	0.297	0.767
Spatial – Planar	1.33%	0.851	0.395
Move Self – Move Object	2.32%	1.393	0.164
Falsifying – Verifying	3.15%	1.593	0.111
SOT			
Holistic – Analytic	1.51%	0.922	0.357
Spatial – Planar	0.77%	0.554	0.581
Move Self – Move Object	0.76%	0.500	0.617
Falsifying – Verifying	0.25%	0.197	0.843

from the pretests to the posttests compared to the other three project groups B, C, and D (-summarized to one group for this investigation -)? The multivariate consideration shows no effects between both groups in the 3DW-test (p = 0.084), in the MRT (p = 0.223), and in the SOT (p = 0.527). In the DAT a global statistical significant effect can be seen. (Pillai-trace = 0.016; $F_{4;681}$= 2.808; p = 0.025). This significant effect can be attributed to the use of the falsifying strategy of group A ($F_{1;684}$ = 7.74; p = 0.006). The students of group A used the falsifying strategy in the posttests (difference to the pretests: -0.68) more often, whereas the other three groups show a tendency towards the verifying strategy (difference to the pretests: +0.13).

5.7. Do promising strategies for assessing spatial tasks exist?

To find out more about the influence of the spatial task solution strategies used by the students in their performance, we used regression models in which gender, school type, school level and all the items of the strategy questions were included. In the 3DW-test the variance explanation increases from 16.3% to 18.0%, so the strategies contributed significantly (F = 4.73; p = 0,001), most of all the two strategies analytic strategy (ß= 0.096; p = 0.005) and the spatial thinking strategy (ß= 0.078; p = 0.025) [49]. In the DAT the variance explanation increases from 11.8% to 15.6% (F = 9.48; p <0.001) if we include the spatial task solution strategies. Just as in the 3DW-tests the same two strategies contributed to an increase in performance: The analytic strategy (ß= 0.109; p = 0.002) and the spatial thinking strategy (ß= 0,169; p <0,001). In the MRT the variance explanation increases from 16.4% to 23.7%

($F = 18.68$; $p < 0.001$). Three strategies contributed significantly: The holistic strategy (ß= 0.115; p = 0.001) in contrast to the analytic strategy, the spatial strategy (ß= 0.202; p <0.001) in contrast to the planar strategy and the move object strategy (ß= 0.098; p = 0.003) in contrast to the move self strategy. Only the falsifying strategy does not contribute significantly (p = 0,438). Finally, in the SOT the variance explanation increases from 15.7% to 18.7% ($F = 7.58$; $p < 0.001$), caused by the move object strategy (ß= 0.093; p = 0.007) and the falsifying strategy (ß= 0.150; p <0.001) [49].

6. Discussion and Prospects

903 students (between 12 and 14 years old) took part in the research project GeodiKon in 2013 and 2014. In the pretests and posttests many data were collected allowing a wide variety of analyses regarding the development of spatial ability. Both research questions (see chapter 2; Training of each factor of spatial ability and training of the repertoire of strategies for assessing spatial tasks will lead to an improvement of spatial ability) cannot be answered in a simple and direct way, because the three treatment groups A, B, and C had a very similar improvement in performance (Table 2 and Figure 7). Many factors can be made responsible for this: ripening process effects, development process effects, influencing effects from school and from out of school, maybe also the short treatment period (12 learning weeks), and the short time duration per week (about 25 minutes per week during the geometry lessons). Stretching of the treatment phase (up to one year) and stretching of the time duration per week spent for the project (up to one whole lesson) could be a way to make significantly different developments visible in the three treatment groups. Independent of that it could make sense to have another posttest after one or two years for all the probands. The results could show long term effects.

It is remarkable that even during the very short treatment phase of 12 weeks the students showed a highly significant increase of performance in all four spatial ability tests (Table 4). Many effects can be responsible for this unspecific trend: learning effects [50], training effects because of the test repetition, ripening process effects, development process effects, treatment effects, and combinations of these effects [13]. The highly significant increase of performance could be a verification of the research work of Thurstone (1955) and Bloom (1971). They argued that children between 5 and 14 years of age show a very high potential for the development of their spatial ability. This leads to the conclusion that we should make as much effort as possible to train, support, and encourage spatial ability in school from the very beginning (age of 5 or 6 years) up to 14 years.

We combined the three treatment groups A, B, and C to one so-called GZ-group. All members of this new group have in common that they all have the subject "Geometrisches Zeichnen" (GZ) in their curricula. So this fact enables us to analyse many effects on students who have geometry lessons compared to students who have no geometry lessons (= control group). It can be noted that the GZ-group performs much better in every of the four spatial ability tests used in the project than the control group. In two tests (3DW-test and MRT) the students of the GZ-group have a significantly higher performance than the probands of the control group even in the pretests (Table 4). The increase of performance of the GZ-group is significantly higher during the treatment phase than in the control group, especially in the MRT. So the MRT is a very sensitive test. What are the reasons for this? Is it because the MRT is a very good and sensitive marker for the subject GZ, or is it (– the other way round –) because the subject GZ trains mainly the factor mental rotation, for which factor the MRT is a marker? It has to be noted that the four spatial ability tests which were used in the project GeodiKon (3DW-test, DAT, MRT and SOT) are "classical" paper-pencil-tests. These tests are apt to show the abilities of the students in the four factors of spatial ability visualization, spatial relations, mental rotation, and spatial orientation. Other spatial abilities (e.g. dynamic spatial ability, small scale/large scale spatial ability, and working memory), which have mainly been identified during the last 20 years, were not in the focus of this project. This leads to the following questions: Which kind of spatial abilities do we train in school? Mainly "classical" spatial abilities, or even also the "new" spatial abilities as mentioned above? Which kind of spatial abilities do we train especially in geometry lessons? Should we include more training of "new" spatial abilities (e.g., dynamics, small/large scale, working memory, moving in three dimensions)? Further projects will pay attention to these questions.

The analysis of the class effect and the teacher effect points out that in the 3DW-test, in the DAT, and in the MRT a substantial class effect is recognizable, which means that students from one class do have a more homogenous increase of performance than individuals from the whole sample. In contrast to this there is no class effect recognizable in the SOT. Does this allow the conclusion that the SOT measures individual performances, and that especially spatial orientation is a very individual ability? Does this also indicate that teachers matter [21], that teachers have a very special and responsibly key role in class?

The gender sensitive analyses of the project show that female students have a significant treatment effect in the 3DW-test. It is remarkable that in all three treatment groups the increase of performance in the 3DW-test is much higher for girls than for boys, and that it is exactly the other way round in the control group. Here the male probands have a higher improvement than the female probands (Table 5 and Figure 10). In the MRT and the SOT boys do have a higher increase of performance than the girls. What are the reasons for this effect? Does it show that girls do have a better development of the factor visualization than the boys? Was the special treatment a reason for this improvement?

The MRT was the only speeded-power test in the test battery of the project. Male probands edited more examples and they also had more items correctly solved than female students. In the SOT male students had a better performance in the pretests and in the posttests.

The gender sensitive analyses point out that male individuals have better spatial abilities only in the two factors mental rotation (- and here maybe only if tested with the help of a speeded-power test -) and spatial orientation. This conclusion confirms research results of Newcombe et al. [42], Glück et al. [16] and Hosenfeld et al. [25].

Individuals use a big variety of different strategies for solving spatial ability tasks, and they even combine them in many different ways. This leads to the conclusion that students should be familiar with a big repertoire of different solving strategies for spatial ability tasks to be able to use them in many different ways and combinations. Students must develop a kind of metaknowledge to be able to handle this wide repertoire consciously [27]. Students very often change their strategies between the pretest to the posttest in same tasks. This is an indication for the fact that with growing routine individuals edit tasks in a different way. Individuals use more new and efficient strategies only when they have sufficient routine in a topic. This leads to the following didactical guiding idea: Teachers should discuss special and selected topics long enough that students can develop a sufficient routine in these fields. Only then students will get to know new and efficient solving strategies even in school and learn how to use them in a meaningful way.

Acknowledgments

The research project GeodiKon was financed by the Federal Ministry for Education, Arts and Culture (BMUKK 20.040/0012-I/7/2012) and the University of Education of Salzburg. The project was also supported by the University of Education of Lower Austria and Vienna, the University of Education of Styria, the University of Innsbruck, the University of Salzburg, theUniversity of Vienna, and the Technical University of Vienna. The workgroup for didactical innovations (ADI) analyses the crosslinks between factors and strategies and the didactic implications for the teaching of geometry. We thank Erich Svecnik (bifie) for the statistical analyses and Georg Gittler (University of Vienna) for the cooperation regarding to the 3DW-test and the anonymous referees.

References

[1] M. Baenninger, N. Newcombe: *The role of experience in spatial test performance: a meta-analysis*. Sex Roles **20**, 327–343 (1989).

[2] E.S. Barratt: *An analysis of verbal reports of solving spatial problems as an aid in defining spatial factors*. J. Psychol. **36**, 17–25 (1953).

[3] G.K. Bennett, H.G. Seashore, A.G. Wesman: *Differential aptitude tests, forms S and T*. The Psychological Corporation, New York 1973.

[4] Bifie (Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens): *Kompetenzen und Modelle*. Retrieved from `https://www.bifie.at/node/49`.

[5] B.S. Bloom: *Stabilität und Veränderung menschlicher Merkmale*. Transl. by D. Eggert, G. Eggert, Beltz, Weinheim, Berlin und Basel 1971.

[6] D. Bryant, B. Tversky: *Mental representations of perspective and spatial relations from diagrams and models*. J. Exp. Psychol.-Learn. Mem. Cogn. **25**, 137–156 (1999).

[7] J.B. Carroll: *Human cognitive abilities. A survey of factor-analytical studies*. Cambridge University Press, New York 1993.

[8] R.B. Cattell: *Theory of Fluid and Crystallized Intelligence: A Critical Experiment*. J. Educ. Psychol. **54** (1963).

[9] A. Dünser: *Trainierbarkeit der Raumvorstellung mit Augmented Reality*. PhD Thesis, University of Vienna, Department of Psychology, 2005.

[10] A.A.H. El Koussy: *The visual perception of space*. Br. J. Psychol. **20**, 1–80 (1935).

[11] J.W. French: *The description of aptitude and achievement tests in terms of rotated factors*. University of Chicago Press, Chicago 1951.

[12] H. Gardner: *The Unschooled Mind: How Children Think and How Schools Should Teach*. Basic Books, New York 1991.

[13] G. Gittler: *GeodiKon: Analysen und Ergebnisse der 3DW-Daten*. Retrieved from `http://www.geotic.at/docs/Gittler-Auswertung-3DW-GEODIKON.pdf`

[14] G. Gittler: *Entwicklung und Erprobung eines neuen Testinstruments zur Messung des räumlichen Vorstellungsvermögens*. Z. Diff. Diagn. Psychol. **5**/2 (1984).

[15] J. Glück, O. Vitouch: *Psychologie*. In S. Günzel (Hrsg.), *Raumwissenschaften*, Suhrkamp, Frankfurt am Main 2009, pp. 324–337.

[16] J. Glück, H. Kaufmann, A. Dünser, K. Steinbügl: *Geometrie und Raumvorstellung — Psychologische Perspektiven*. Informationsblätter der Geometrie (IBDG) **24**/1, 4–11 (2005).

[17] M. Grüssing: *Wieviel Raumvorstellung braucht man für Raumvorstellungsaufgaben? Strategien von Grundschulkindern bei der Bewältigung räumlich-geometrischer Anforderungen*. Zentralblatt für Didaktik der Mathematik (ZDM) **34**/2 (2002).

[18] J.P. Guilford: *The Nature of Human Intelligence*. McGraw-Hill, New York 1967.

[19] J.P. Guilford: *The structure of intellect*. Psychol. Bull. **53**/4, 267–293 (1956).

[20] J.P. Guilford, W.S. Zimmerman: *The Guilford-Zimmerman Aptitude Survey*. J. Appl. Psychol. **32**, 24–34 (1948).

[21] J. Hattie: *Lernen sichtbar machen* (= überarbeitete deutschsprachige Ausgabe von W. Beywl, K. Zierer: *Visible Learning*), Schneider Verlag GmbH., Hohengehren 2013.

[22] M. Hegarty, D. Waller: *Individual differences in spatial abilities*. In P. Shah, A. Miyake (eds.): *The Cambridge Handbook of Visuospatial Thinking*, Cambridge University Press, Cambridge 2005, pp. 121–169.

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

[24] D.L. Hintzman, C.S. O'Dell, D.R. Arndt: *Orientation in cognitive maps*. In Cognitive Psychol. **13**, 149–206 (1981).

[25] I. Hosenfeld, B. Strauss, O. Köller: *Geschlechtsdifferenzen bei Raumvorstellungsaufgaben – eine Frage der Strategie?* Z. Pädagog. Psychol. **11**/2 (1997).

[26] M.A. Just, P.A. Carpenter: *Cognitive Coordinate Systems: Accounts of Mental Rotation and Individual Differences in Spatial Ability*. Psychological Review **92** (1985).

[27] H. Kaufmann: *Lösung- und Bearbeitungsstrategien bei Raumvorstellungsaufgaben*. In Raumgeometrie — intuitiv und konstruktiv, CD, 2008.

[28] P. Kline: *The handbook of psychological testing*, 2nd ed., Routledge, London 1999.

[29] M. Kozhevnikov, M. Hegarty: *A dissociation between object-manipulation spatial ability and spatial orientation abilities*. Mem. Cognition **29**, 745–756 (2001).

[30] M. Kraker, A. Asperl, M. Forstner, D. Hochhauser, M. Husty, M. Liebscher, G. Maresch, D. Miestinger, O. Röschel, E. Wurm: *Die kompetenzorientierte Reifeprüfung. Darstellende Geometrie. Richtlinien und Beispiele für Themenpool und Prüfungsaufgaben.*. bmukk, Wien 2012.

[31] M. Liebscher, G. Breyer, S. F$\tilde{\text{A}}\frac{1}{4}$rst H. Heugl M. Kraker, C. Preis, E. Svecnik, I. Liegl, G. Plattner: *Praxishandbuch Mathematik AHS Oberstufe. Auf dem Weg zur standardisierten kompetenzorientierten Reifeprüfung. Teil 1*. Vienna: Bifie (Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens), 2013.

[32] M.C. Linn, A.C. Petersen: *Emergence and characterization of sex differences on spatial ability: a meta-analysis*. Child Dev. **56**, 1479–1498 (1985).

[33] D.F. Lohman: *Spatial abilities: A review and re-analysis of the correlational literature*. Techn. Rep. No. 8, Aptitude Research project, Stanford University 1979.

[34] D.F. Lohman: *Spatial abilities as traits, processes, and knowledge*. In R.J. Sternberg (ed.): *Advances in the psychology of human intelligence* **40**, Erlbaum, Hillsdale/NJ 1988, pp. 181–248.

[35] H.P. Maier: *Räumliches Vorstellungsvermögen: Komponenten, geschlechtsspezifische Differenzen, Relevanz, Entwicklung und Realisierung in der Realschule*. Europäische Hochschulschriften: Reihe 6, Psychologie, Band 493, 1994.

[36] P. Mair, R. Hatzinger, M.J. Maier: *eRm: Extended Rasch Modeling. 0.15-5*. Retrieved from `http://erm.r-forge.r-project.org/`, 2015.

[37] G. Maresch: *Spatial Ability — The Phases of Spatial Ability Research*. J. Geometry Graphics **17**/2, 237–250 (2013).

[38] G. Maresch, T. Müller, K. Scheiber (eds): *GeodiKon. Die Lernmaterialien. Praktische Raumvorstellungsübungen für den Geometrie- und Mathematikunterricht mit Lösungen*. Studienverlag, Innsbruck 2014.

[39] G. Maresch: *Raumintelligenz – Die Phasen der Raumintelligenzforschung*. Informationsblätter der Geometrie (IBDG) **32**/1 (2013).

[40] M.G. McGee: *Human spatial abilities: psychometric studies and environmental, genetic, hormonal and neurological influences*. Psychol. Bull. **86**/5, 889–918 (1979).

[41] S. Mick, S. Eibl, J. Gabl, D. Hochhauser, S. Ranger, J. Schmied: *Arbeitsunterlagen zu einem kompetenzorientierten Unterricht aus Geometrischem Zeichnen*. Retrieved from `http://www.geometriekompetenzen.at`

[42] N.S. Newcombe, L. Mathason, M. Terlecki: *Maximization of spatial competence: more important than finding the cause of sex differences*. In A. McGillicudy-De Lisi, R. De Lisi (eds.): *Biology, Society, and Behavior: The Development of Sex Differences in Cognition.* Ablex, Wesport/CT 2002, pp. 183–206.

[43] M. Peters, B. Laeng, K. Latham, M. Jackson, R. Zaiyouna, C. Richardson: *A Redrawn VandjnmRRREDenberg & Kuse Mental Rotations Test: Different Versions and Factors that affect Performance*. Brain Cognition **28**, 39–58 (1995).

[44] G. Rasch: *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research, Copenhagen 1960, expanded ed. with foreword and afterword by B.D. Wright. The University of Chicago Press, Chicago 1980.

[45] D.H. Rost: *Raumvorstellung. Psychologische und pädagogische Aspekte*. Beltz, Weinheim 1977.

[46] K. Schultz: *The contribution of solution strategy to spatial performance*. Can. J. Psychology **45** (1991).

[47] E. Souvignier: *Förderung räumlicher Fähigkeiten – Trainingsstudien mit lernbeeinträchtigten Schülern*. Pädagogische Psychologie und Entwicklungspsychologie **22** (2000).

[48] C. Spearman: *General Intelligence, objectively determined and measured.* Am. J. Psychol. **15**, 201–293 (1904).

[49] E. Svecnik: *GeodiKon Pretest – Ergebnisse der Datenanalysen. Statistische Auswertung der Pretests.* Retrieved from `http://www.geotic.at/docs/Pretest-Analysen-GeodiKon-Svecnik-2013-12.pdf`, 2013.

[50] E. Svecnik: *GeodiKon Posttest – Ergebnisse der Datenanalysen. Statistische Gesamtauswertung des Projekts GeodiKon.* Retrieved from `http://www.geotic.at/docs/Posttest-Analysen-GeodiKon-Svecnik-140703.pdf`, 2014.

[51] E. Svecnik: *GeodiKon-Rasch-Analyses*. Retrieved from `http://www.geotic.at/docs/Svecnik-Rasch-Analysis-3DW-DAT-MRT.pdf`, 2015.

[52] E. Svecnik: *GeodiKon-Aposteriori-Students-t-Tests*. Retrieved from `http://www.geotic.at/docs/Svecnik-Aposteriori-t-test.pdf`, 2015.

[53] B.G. Tabachnick, L.S. Fidell: *Using Multivariate Statistics* (5th ed.), Pearson, Boston 2007.

[54] L.L. Thurstone: *Primary Mental Abilities*. The University of Chicago Press, Chicago 1938.

[55] L.L. Thurstone: *Some primary abilities in visual thinking*. In Psychometric Laboratory Research Report No. 59. University of Chicago Press, Chicago 1950.

[56] P.E. Vernon: *The structure of human abilities (Rev. ed.)*. Methuen, London 1961.

[57] D. Voyer, S. Voyer, P. Bryden: *Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables*. Psychol. Bull. **117**, 250–270 (1995).

[58] B.T. West, K.B. Welch, A.T. Galecki: *Linear Mixed Models: A Practical Guide Using Statistical Software*. Chapman & Hall/CRC, New York 2007.

Received December 5, 2014; final form June 3, 2015