

# R-Cube-SR Test

## A New Test for Spatial Relations Distinguishable From Visualization

Benedict C. O. F. Fehrer

Department of Psychology of Education, University of Mannheim, Germany

**Abstract:** Visualization and spatial relations (mental rotation) are two important factors of spatial thinking. Visualization refers to complex visual-spatial transformations, whereas spatial relations refer to simple mental rotation of visualized objects. Conventional spatial relations tests, however, have been found to be highly correlated with visualization tests because solving items through mental rotation might involve visualization ability due to the complexity of the visual materials of these tests. In two studies ( $N = 51$ ,  $N = 109$ ), a new computer-based test for spatial relations, the R-Cube-SR Test, was developed and validated. The R-Cube-SR Test utilizes simple, single-colored cubes as rotated visual materials. Reliability estimates of the reaction times reach  $\omega = .87$ . Correlations with standard tests of spatial relations (up to  $r = .55$ ) were significantly higher than with visualization tests, such as the new R-Cube-Vis Test (Fehrer, 2020), which uses the same visual materials. This was supported by CFAs. It is concluded that the new R-Cube-SR Test is a valid measure of spatial relations. Both tests together, the R-Cube-Vis and R-Cube-SR, as specific tests for their respective factor, now, are able to provide a differential diagnosis of a participant's spatial thinking ability using the same visual materials.

**Keywords:** spatial thinking, spatial ability, spatial relations, computer-based testing, Rubik's cube

Spatial thinking is a nonverbal intelligence component (e.g., Paivio, 2014). Especially in developing specialized knowledge in STEM (science, technology, engineering, and mathematics) domains, spatial thinking is important (Wai et al., 2009). Furthermore, it is related to imagery preferences and creativity (Kozhevnikov et al., 2013), but it is also related to vocational interests (Pässler et al., 2015). Spatial thinking ability plays a critical role in how people benefit from supporting aids in learning from visualizations (Münzer, 2012) and choosing solving strategies in spatial tasks (Wang & Carr, 2014). The heterogeneous construct of spatial thinking consists of different factors. For each of these factors, there exist specific tests (see Hegarty & Waller, 2005). However, the standard tests of the two main factors, visualization, and spatial relations, are highly correlated (e.g., Kozhevnikov et al., 2002; Kozhevnikov & Hegarty, 2001; Miyake et al., 2001). One reason might be that the stimulus materials of the spatial relations tests are too complex to measure the pure ability of spatial relations and, therefore, visualization ability is needed to solve the items.

The goal of the present study was to develop a new test for spatial relations that uses sufficiently simple stimuli to measure pure spatial relations ability. The materials of the new test, the R-Cube-SR Test, are very similar to the materials utilized in the R-Cube-Vis Test, a new test for the visualization factor (Fehrer, 2020). This will allow simple comparisons between both abilities. The testing materials were validated regarding construct and criterion validity.

## Spatial Thinking Measures

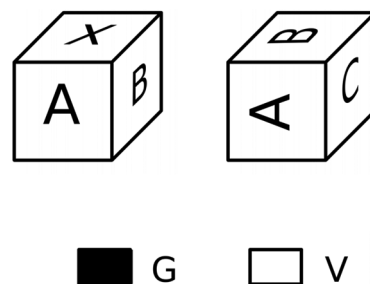
Based on existing spatial thinking tests, factor-analytic studies attempt to differentiate spatial thinking according to different aspects such as manipulating or rotating an object (see Hegarty & Waller, 2005). Since spatial thinking is defined differently from tight to broad, authors describe two to five different factors (e.g., Carroll, 1993; McGee, 1979). The present work bases on Carroll (1993) who reanalyzed more than 90 studies. He found a five-factor structure that could also be confirmed by Burton and Fogarty (2003). The present study focuses on the first two main factors, visualization and spatial relations.

In his definitions, Carroll (1993) refers to Ekstrom et al. (1976b): Visualization means “the ability to manipulate or transform the image of spatial patterns into other arrangements” (Ekstrom et al., 1976b, p. 173). Spatial relations is “the ability to perceive spatial patterns or to maintain orientation with respect to objects in space” (Ekstrom et al., 1976b, p. 149). In contrast to visualization, where the configuration of parts of the object is changed, spatial relations means manipulating the whole object (Ekstrom et al., 1976b). According to Ekstrom et al. (1976b), tests of visualization are also less speeded and use more complex objects than tests of spatial relations. Pellegrino et al. (1984) interpreted visualization and spatial relations as end points of a continuum with speeded tests consisting of simple items on the spatial relations side and power tests

(less speeded tests) consisting of complex items on the visualization side. Tests in the middle of the continuum are thought to consist of medium complex tasks. Therefore, standard tests for visualization are constructed as power tests and use accuracy as a measure (e.g., the Paper Folding Test, PFT; Ekstrom et al., 1976a), whereas spatial relations tests are constructed as speeded tests and use reaction times, if possible, as the preferred measure (e.g., a test using simple, two-dimensional figures obtaining a chronometric measure for mental rotation speed, Jansen-Osmann & Heil, 2007, termed here as the Chronometric Test, CT). However, some conventional spatial relations tests exist as paper/pencil versions only. Therefore, accuracy is measured (e.g., Cube Comparison test, CC; Ekstrom et al., 1976a). A test from the middle of the continuum is the Mental Rotation Test (MRT, Vandenberg & Kuse, 1978). The MRT presents more complex three-dimensional figures such as visualization tests but demands mental rotations of these figures such as spatial relations tests.

Although visualization and spatial relations can be seen as placed at the two opposite ends of this continuum (Pellegrino et al., 1984), the correlation patterns based on the existing, conventional tests provide only weak support for the distinguishability of both spatial thinking factors. The correlations between different tests measuring spatial relations vary between  $.34 \leq r \leq .62$  (see Hegarty & Waller, 2004; Kozhevnikov et al., 2002; Kozhevnikov & Hegarty, 2001; Vandenberg & Kuse, 1978). Correlations between different visualization tests can only be approximately estimated by the correlation between PFT and MRT,  $.51 \leq r \leq .63$  (Blajenkova et al., 2006). However, correlations between spatial relations and visualization tests are in the same range,  $.40 \leq r \leq .57$  (see Kozhevnikov et al., 2002; Kozhevnikov & Hegarty, 2001; Miyake et al., 2001).

The high correlations between both spatial thinking factors that approach the correlations between tests of the same factor can be problematic in research focusing on differences between both factors. These correlation patterns might be caused either by an existing strong similarity of spatial relations and visualization or by an invalid measurement of at least one of the two factors. Assuming that all tests of one factor contain the same “invalid” aspect(s), this might cause a misinterpretation of the specific tests. One reason for such a misinterpretation might be the factor-analytic approach that led to these factors. Factor analyses are able to group tests but are unable to validate these grouped tests with respect to a certain construct. Especially for the spatial relations tests, this might be problematic. Although the most spatial relations tests fulfill the definition of spatial relations (i.e., participants have to rotate a visual object to solve an item), the items might be too complex for the usage in speed tests. For example, Münzer et al. (2018) found an overall accuracy of the CT of around 70% in a stu-



**Figure 1.** Cube Comparison Test (CC). G = gleich (same); V = verschieden (different). Here, the correct answer is “G”.

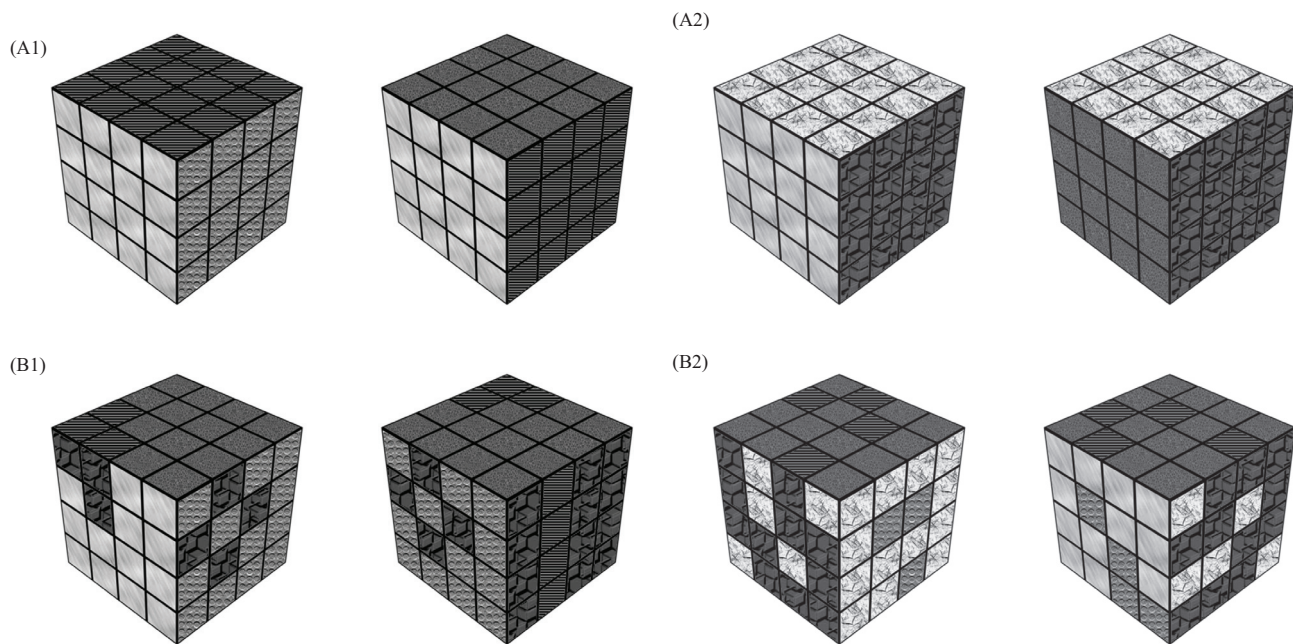
dent population, and Henn et al. (2018) found an overall accuracy of around 50% for the CC in a sample of surgical trainees. Especially for the CT, this low accuracy rate is problematic considering that the reaction times are based only on the correctly solved items as well as the recommendation to exclude participants with an error rate of 30% and higher (Jansen et al., 2013). The items of the CC might also be too complex. Each item of the CC presents two cubes with letters on their sides. Participants have to decide whether both cubes are equal except for rotation (Figure 1).

Although this task seems to fulfill the spatial relations’ definition, the participants also have to check the orientation of the letters to solve the task. This mental transformation of the letters’ orientation might also demand visualization ability. Such complex elements in the items of spatial relations tests might be one reason for the partly strong positive correlations between spatial relations and visualization tests. Hence, a new spatial relations test with simpler tasks is needed.

The newly developed test for spatial relations, the R-Cube-SR Test, was intended to present simple items with the goal that nearly all participants should solve sufficiently many items to maintain a pure(*r*) measure for spatial relations. Similar to Jansen et al. (2013), the maximal error rate per participant should not exceed 30%. Additionally, the second set of more complex items, similar to existing tests, was also considered for comparison reasons. Both item sets of the R-Cube-SR Test were analyzed in Study 1 with respect to the standards for evidence of validity formulated by the American Educational Research Association et al. (2014). A second study (Study 2) was conducted to replicate the results of Study 1 in a larger sample.

## Stimulus Materials of the R-Cube-SR Test

The items of the R-Cube-SR Test show two cubes, one on each side, similar to Rubik’s cubes (Figure 2). Both cubes are colored with six colors blue, brown, green, red, white, and yellow. Since only three sides are visible, the participants might see less colors. The cube on the left side of



**Figure 2.** Sample items of the R-Cube-SR Test (here, only in grayscale, the actually used figures can be found in the supplementary material, Fehring, 2021b): (A1) plain and possible item; (A2) plain, impossible; (B1) patterned, possible; (B2) patterned, impossible.

each item is the target figure. The cube on the right side of each item is a rotated cube. Participants have to decide whether it is *possible* or *impossible* to rotate the left cube to obtain the right cube. The R-Cube-SR Test was created in two versions (*Plain* and *Pattern*). The items of the Plain-version are plain colored cubes and are simpler than the items of the standard spatial relations tests (Figures 2A1 and 2A2). Cube sides of standard spatial relations test are usually more complex by showing patterns or arrangements of dots or letters (see Figure 1). These visual symbols that have, therefore, also to be checked for their correct orientation after the cube was rotated which involve additional cognitive processes. Therefore, the plain version of the R-Cube-SR Test is simpler than the existing cube comparison tests. Hence, it was assumed that these items could be utilized to measure spatial relations in a pure form. Furthermore, the second set of items, the Pattern-version, with more complex items was created to compare the Plain-version and the standard tests. Items of the Pattern-version show patterned cubes (Figures 2B1 and 2B2) that are more complex and more similar to the items of the conventional tests.

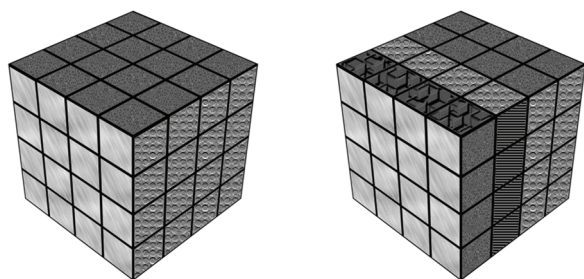
The R-Cube-SR Test (Plain) is constructed as a speeded test (Pellegrino et al., 1984). The logarithmized reaction times (RT) averaged over all possible, correctly answered items are utilized as the relevant measure, such as in the CT. Items of the Pattern-version of the R-Cube-SR Test were expected to be more difficult to solve than in the Plain-version and comparable to the items of the CC test.

Therefore, the accuracy over all items (ACC) was analyzed as additional measure of the Pattern-version, additional to the logarithmized reaction times averaged over all possible items.

The item set of the R-Cube-SR Test is built in the same manner and use the same figures as another spatial thinking test for the factor visualization, the R-Cube-Vis Test for visualization (Fehring, 2020). However, in the R-Cube-Vis Test, the cubes have not to be rotated as a whole, but single elements of a cube have to be manipulated mentally in succession, as with a Rubik's cube, to (possibly) obtain the appearance of the comparison cube (Figure 3). Therefore, both tests combined, R-Cube-SR and R-Cube-Vis, can assess different spatial factors using the same stimulus materials, which is not possible with the single conventional tests.

Rational and theoretical reasons can be derived to provide evidence-based for validation on test content that "can be obtained from an analysis of the relationship between the content of a test and the construct it is intended to measure." (American Educational Research Association et al., 2014, p. 14). This evidence can be confirmed for the R-Cube-SR Test (both versions) regarding the definition of spatial relations (see above). In order to rotate the whole cube of an R-Cube-SR item, participants have to "perceive spatial patterns" (one side of the cube) "or maintaining orientation with respect to objects in space" (the position of one side with respect to the other two sides). In this way, the items together with the test instruction correspond to the definition of spatial relations.





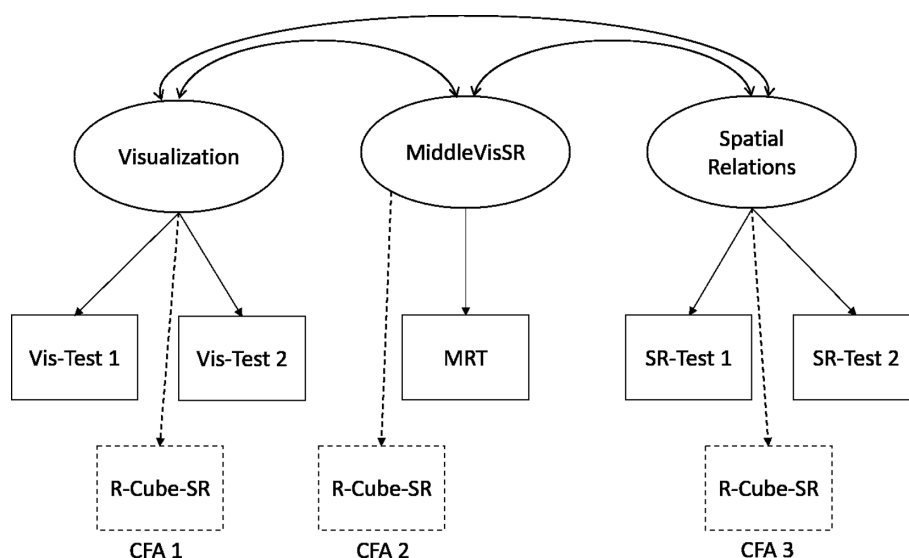
**Figure 3.** Sample item of the R-Cube-Vis Test of a medium difficulty level (here, only in grayscale, the actually used figures can be found in the supplementary material, Fehringer, 2021b).

## Study 1: Validation of the R-Cube-SR Stimulus Materials

The stimulus materials of both versions of the R-Cube-SR Test were validated with respect to evidence-based on relations to other variables, including the relationship to other tests measuring the same or a different construct as well as to external criteria. This validity aspect refers to “external variables [that] may include measures of some criteria that the test is expected to predict, as well as relationships to other tests hypothesized to measure the same constructs, and tests measuring related or different constructs.” (American Educational Research Association et al., 2014, p. 16) It was expected that the correlations with two standard tests for spatial relations are significantly positive (evidence for convergent validity) and that the correlation with two visualization tests should be lower (discriminant evi-

dence of validity). The size of the correlation with a further test from the middle of the continuum between spatial relations and visualizations (the MRT) should be between the correlations with the spatial relations and the visualization test. The pairwise correlations between the R-Cube-SR test and the spatial relations tests should be at least at the minimum size of correlations between spatial relations tests based on the current literature ( $r = .34$ , see above) and should be comparable to the correlations between both conducted spatial relations tests. Hence, the sample size should be at least  $N = 49$  with a directed hypothesis, an  $\alpha$ -level at .05, and a test power at .80 (recommended by Cohen, 1988) according to the software G\*Power (Version 3.1.9.2, Faul et al., 2007). Additionally, three confirmatory factor analyses (CFA) were conducted with three latent variables: *Visualization* representing the visualization tests, *Spatial Relations* representing the spatial relations tests, and a latent variable *MiddleVisSR* for the MRT as in between both ends of the continuum. The three CFAs differ with regard to which latent variable the R-Cube-SR Test is assigned to (Figure 4). It was expected that the CFA with the R-Cube-SR Test as manifest variable of the latent *Spatial Relations* variable would best fit the data. This would support the R-Cube-SR Test as a spatial relations test.

Furthermore, evidence for the validity regarding relationships with external criteria was also considered for school grades in the native language and Mathematics. There was no specific expectation for the effect sizes, because the correlations with spatial relations with mathematical and verbal abilities range from  $.20 \leq r \leq .52$  and  $.17 \leq r \leq .65$  (Gunderson et al., 2012; Hegarty & Kozhevnikov,



**Figure 4.** Schematic representations of the three conducted confirmatory factor analyses (CFAs) differing with regard to which latent variable the R-Cube-SR Test was assigned to.

1999). However, it was expected that the correlations of the school grades with the R-Cube-SR should be similar to the correlations of the school grades with the standard spatial relations tests in the present studies. Both presented studies were also included in the validation of the R-Cube-Vis Test. The results concerning the R-Cube-Vis Test are presented in Fehringer (2020).

## Method

### Participants

The study was conducted with a sample of 51 participants (39 female, 12 male) from a German University. For their participation, they received course credit. On average, the participants were 21 years of age ( $M = 21.14$ ,  $SD = 2.17$ ), with the youngest being 17 and the oldest 26. Two further participants were discarded for all analyses because of a higher error rate than 30% in the target test, the R-Cube-SR Test (Plain).

### Materials

Both item versions of the R-Cube-SR Test, Plain and Pattern (Figure 2), consisted of 12 possible and 12 impossible items. The order of the version was fixed beginning with Plain. The items within each version were randomly presented. Before each block, there was one trial phase with two possible and two impossible items that could be repeated by the participant. However, a technical error led to a wrong presentation of one possible plain item. Hence, the Plain version of the R-Cube-SR Test consisted of eleven possible and 13 impossible items.

The performed visualization tests were the Paper Folding Test (PFT) and the R-Cube-Vis Test. The PFT consists of 20 items presented on two sheets with 10 items each. Each item shows a quadratic paper sheet that is folded and perforated in two to four steps. On the right side of each item, there are five response options whereby each option shows the unfolded paper sheet, but with different patterns of holes. The task is to decide which option presents the correct pattern of holes resulting from the folding and perforating steps. The participants had 3 min for each page and got 1 and 0 points for correctly and incorrectly answered items.

The R-Cube-Vis Test was presented in its long version with 24 possible and 24 impossible items for six difficulty levels. After each item, the participants were asked for their confidence on a 4-point-scale. The weighted accuracy (wACC), a combined measure based on the accuracy over all possible items and the respective confidence rating, was considered for analyses (Fehringer, 2020).

As tests for spatial relations, the Chronometric Test (CT; Jansen-Osmann & Heil, 2007), and the subtest Würfelaufgaben of the Intelligenz-Struktur-Test (IST-WA; Amthauer

et al., 1999) were conducted. The CT consists of 60 items presenting two-dimensional figures ("primary mental ability figures", Thurstone, 1958). In each item, participants have to decide whether both objects are equal except for rotation. There are 16 items, 8 same and 8 mirrored, for 45°, 90°, 180° disparity angles. Twelve additional items have a disparity angle of 0° and are all mirrored. The CT was computer-based and uses reaction times for possible, correctly solved items as the measure.

The IST-WA shows cubes with dots, squares, and triangles on their sides. The test consists of two blocks. The participants have to assign to each cube of a target set (10 cubes per block) the specific cube of a given set (five cubes) that shows its rotated version. For each correctly assigned cube from the target set, the participants got 1 point, otherwise 0 points. The IST-WA is similar to the CC as a standard spatial relations test and was repeatedly validated as part of an intelligence test.

Furthermore, the Mental Rotation Test (MRT; Vandenberg & Kuse, 1978) was conducted from the middle of the continuum between visualization and spatial relations. The MRT utilizes complex three-dimensional objects as figures in its stimulus materials. Each task shows one object on the left side and four objects on the right side. Two of these four objects are equal to the object from the left side but rotated. The other two objects are different. Participants have to detect both equal objects on the right side. The MRT consists of 24 items, which were presented on two sheets with twelve items on each. The participants had 4 min per sheet and got 1 point, if the two equal figures were correctly detected; otherwise, the participants got 0 points. In the current study, a redrawn version, the MRT-A (Peters et al., 1995) was used.

Finally, a questionnaire was prepared to ask for participants' age, gender, their experience with the Rubik's Cube and Abitur grades in German and Mathematics. The Abitur is the diploma from German secondary school qualifying for university admission. The grades range from 0.7, that is, the best grade, to 6.0, that is, worst grade.

### Procedure

There were two sessions, A and B, with durations of 90 and 60 min. In session A, participants conducted the PFT and the R-Cube-Vis test in that order. Session B took place one week later. The participants performed the following tests: MRT, IST-WA, R-Cube-SR, CR and the questionnaire (in the indicated order). Both sessions were conducted in an experimental laboratory in groups of up to six participants using partition walls.

### Analyses

The analyses in both studies were conducted using the R statistics (R Core Team, 2017) with the packages plyr

**Table 1.** Descriptive results for each test (Study 1)

	<i>N</i>	<i>M</i>	<i>SD</i>	Range	Skewness	Kurtosis
PFT	50	0.65	0.16	0.25–0.95	−0.52	−0.08
MRT	51	0.61	0.21	0.08–0.96	−0.32	−0.92
IST-WA	51	0.59	0.23	0.15–1.00	0.10	−0.85
CT (ACC)	48	0.93	0.06	0.75–1.00	−0.72	0.10
CT (RT)	48	2,471	602	1,292–3,793	0.43	−0.68
R-Cube-SR, Plain (ACC-all)	51	0.96	0.05	0.79–1.00	−1.22	1.11
R-Cube-SR, Plain (RT)	50	4,709	2,077	2,590–17,031	4.09	21.52
R-Cube-SR, Pattern (ACC-all)	50	0.82	0.13	0.38–1.00	−1.13	−1.03
R-Cube-SR, Pattern (RT)	42	6,649	2,291	2,968–15,935	1.55	4.16
R-Cube-Vis (wACC)	49	3.26	0.39	2.38–3.86	−0.44	−0.95

Note. The different sample sizes result from different numbers of excluded participants for each test. ACC = accuracy over all items; CT = Chronometric Test; IST-WA = Würfelaufgaben of the Intelligenz-Struktur-Test; *M* = mean; MRT = Mental Rotation Test; PFT = Paper Folding Test; RT = reaction time; *SD* = standard deviation; wACC = weighted ACC.

(Wickham, 2011), dplyr (Wickham et al., 2018), reshape (Wickham, 2007), and psych (Revelle, 2017). Reliability was estimated by omega (McDonald, 1999) as factor analytic alternative to Cronbach's  $\alpha$  (Trizano-Hermosilla & Alvarado, 2016) using the package MBESS (Kelley, 2020). The comparisons of the correlations were tested with Fisher's  $z$  utilizing the package cocor (Diedenhofen & Musch, 2015). The multiple imputation was performed with mice (Van Buuren & Groothuis-Oudshoorn, 2011) and miceadds (Robitzsch & Grund, 2020). The confirmatory factor analyses were performed with lavaan (Rosseel, 2012) using a maximum-likelihood estimator. In both studies, the author reported how they determined the sample size, all data exclusions, all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established before data analysis, all measures in the study, and all analyses including all tested models. For the inferential tests, the author reported exact  $p$ -values, effect sizes, and 95% confidence or credible intervals.

## Results

The descriptive results can be found in Table 1. The reaction times showed high values in skewness and kurtosis. Therefore, the values were also logarithmized for the following analyses. The different sample sizes of the tests result from the predefined outlier criteria based on the results of a pre-study, for example, discarding of too few valid answers or pressing only one key. For the further measures using reaction times (CT; the Pattern-version of the R-Cube-SR Test), only participants were included with an error rate of maximal 30%. The lower number of participants of the R-Cube-Vis Test was mainly because of a wrong interpretation of the confidence rating by participants. The pairwise exclusion was applied to maximize the number of participants per correlation. In addition, the obtained correlations were compared with the results of a multiple imputation to check for strong distortions.

First, the correlation patterns among the validation tests will be considered. The highest correlations can be found between both spatial relations tests (IST-WA, CT),  $r = -.55$ ,  $p < .001$ , between both visualization tests (PFT, R-Cube-Vis),  $r = .49$ ,  $p < .001$  and between the PFT, MRT, and IST-WA,  $.46 \leq r \leq .50$ ,  $p < .001$  (Table 2). Therefore, the CT and R-Cube-Vis seem to be the purest tests with descriptively higher correlations with the respective test from the same factor ( $|r| \geq .49$ ,  $p < .001$ ) and smaller correlations with tests from the other factor,  $.22 \leq |r| \leq .37$ ,  $p \leq .142$ . The correlations between CT and IST-WA as well as CT and R-Cube-Vis were significantly different,  $z = 1.83$ ,  $p = .033$ . The correlations between R-Cube-Vis and PFT as well as R-Cube-Vis and CT were marginally significantly different,  $z = 1.45$ ,  $p = .074$ . The results of the multiple imputation were comparable ( $|\Delta r| \leq .02$ ). Significant correlations between the validation tests and the grades in German or Mathematics as well as with the experience with Rubik's cubes could only be found for the PFT with Rubik's cubes' experience ( $r = .33$ ,  $p = .018$ ) and with the grade in Mathematics,  $r = -.30$ ,  $p = .040$ . All results were in accordance with the current literature.

The reliability estimate of the reaction times of both R-Cube-SR versions were good,  $\omega = .84$  [.77; .90] (Plain) and  $\omega = .89$  [.84; .95] (Pattern). The estimate of the accuracy over all items of the Pattern-version was acceptable,  $\omega = .71$  [.59; .84].

In the following, the correlation patterns including the R-Cube-SR Test are examined. As described in the theoretical part, the reaction times were considered as standard measure for a speeded test for both versions. Additionally, the accuracy over all items was also analyzed for the Pattern-version due to the more complex items and, therefore, the lower mean accuracy. The reaction times of both versions of the R-Cube-SR Test showed descriptively the expected pattern with the highest correlations with the CT as a standard test for spatial relations ( $r \geq .48$ ,  $p < .001$ ) and the lowest correlation with the visualization test

**Table 2.** Correlations between the validation tests (Study 1)

	R-Cube-Vis	PFT	MRT	IST-WA
PFT	.49***			
MRT	.36*	.46***		
IST-WA	.28	.48***	.50***	
CT (RT)	-.22	-.37*	-.32*	-.55***

Note. CT = Chronometric Test; IST-WA = Würfelaufgaben of the Intelligenz-Struktur-Test; MRT = Mental Rotation Test; PFT = Paper Folding Test.  
\* $p < .05$ ; \*\*\* $p < .001$ .

R-Cube-Vis,  $r \geq -.04$ ,  $p \geq .791$  (Table 3). These correlation pairs were significantly different,  $z = 2.33$ ,  $p = .010$  (Plain) and  $z = 2.74$ ,  $p = .003$  (Pattern). All other correlations did not differ significantly ( $z \leq -1.16$ ,  $p \geq .123$ ). The descriptively lower correlations between the reaction times of both versions and the second spatial relations test (the IST-WA) were unexpected. One reason might be that the IST-WA is farther away from the spatial relations end of the continuum and, therefore, more in the middle than purer tests of this factor. This is also supported by the descriptively higher correlations of the IST-WA with PFT and MRT compared to the correlations of the CT with the PFT and MRT. The accuracy measure of R-Cube-SR (Pattern) showed lower correlations ( $r \leq .37$ ,  $p \geq .008$ ) and not the expected pattern as a spatial relations test. The results of the multiple imputation were similar to the correlations of the reactions time of the Plain version and of the accuracy of the Pattern version ( $|\Delta r| \leq .02$ ). However, the correlations between the reaction times of the Pattern version were smaller for the PFT ( $r = -.22$ ,  $p = .128$ ), the MRT ( $r = -.06$ ,  $p = .723$ ), and the IST-WA ( $r = -.21$ ,  $p = .161$ ). All correlations were also computed with the female-only subsample for comparison because of the uneven gender proportion. Except for the correlation between the reaction times of R-Cube-SR-Pattern and R-Cube-Vis as well as IST-WA ( $r = .11$ ,  $p = .551$ ;  $r = -.42$ ,  $p = .019$ ), all correlations differed less than ( $|\Delta r| \leq .10$ ). The correlation patterns were the same.

The confirmatory factor analyses (CFAs) were conducted with reaction times of the R-Cube-SR Test of the Plain- and Pattern-version as both showed the expected correlation

**Table 3.** Correlations of the R-Cube-SR Test versions with the validation tests (Study 1)

Measure	R-Cube-Vis	PFT	MRT	IST-WA	CT (RT)
R-Cube-SR-Plain					
RT	-.04	-.33*	-.33*	-.34*	.48***
R-Cube-SR-Pattern					
ACC	.21	.10	.37**	.27	.00
RT	-.02	-.36*	-.31*	-.30	.57***

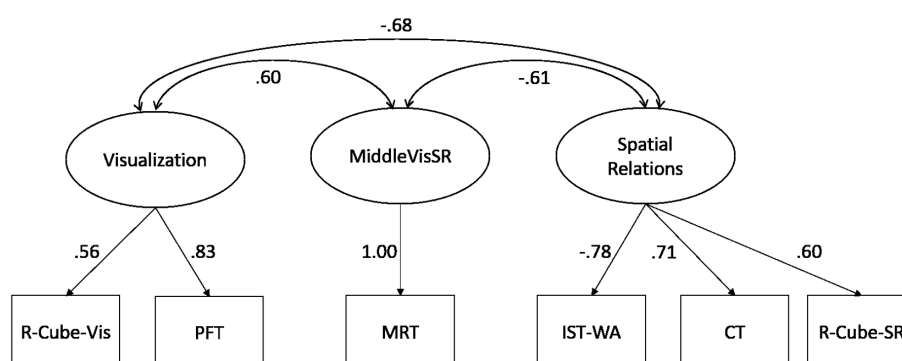
Note. ACC = accuracy over all items; CT = Chronometric Test; IST-WA = Würfelaufgaben of the Intelligenz-Struktur-Test; MRT = Mental Rotation Test; PFT = Paper Folding Test; RT = reaction time. All correlations  $|r| \geq .3$  are bold. \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

patterns. The CFA with the R-Cube-SR Test (Plain) assigned to the latent Spatial Relations variable (CFA 3 in Figure 4 and Figure 5) had the best fit to the data according to the information criteria AIC and BIC ( $AIC[CFA\ 3] = -56.26 < AIC[CFA\ 2] = -53.87 < AIC[CFA\ 1] = -53.18$ ;  $BIC[CFA\ 3] = -30.97 < BIC[CFA\ 1] = -27.89 < BIC[CFA\ 2] = -26.77$ ) and showed also good model fit indices according to Kline (2016),  $RMSEA = .05$  [.00; .20],  $CFI = .99$ ;  $TLI = .97$ . The best fit for the R-Cube-SR Test (Pattern) could also be found, if the test was assigned to the latent Spatial Relations variable according to the information criteria AIC and BIC ( $AIC[CFA\ 3] = -36.35 < AIC[CFA\ 2] = -32.66$ ;  $BIC[CFA\ 3] = -11.05 < BIC[CFA\ 1] = -5.57$ ), whereas the CFA 1 did not even converge. However, the fit indices were only poor to acceptable with  $RMSEA = .13$  [.00; .24],  $CFI = .93$ , and  $TLI = .84$ .

All correlations with grades in Mathematics and German as well as with experience with Rubik's cube were not significant.

## Study 2: Replication

The goal of Study 2 was to confirm the correlation patterns between the R-Cube-SR Test and the validation tests with a larger sample. The Würfelaufgaben (IST-WA) as spatial relations test was replaced by the Cube Comparison Test

**Figure 5.** Confirmatory factor analysis with the R-Cube-SR Test (Plain) assigned to the latent Spatial Relations variable (Study 1).



**Table 4.** Descriptive results for each test (Study 2)

	<i>N</i>	<i>M</i>	<i>SD</i>	Range	Skewness	Kurtosis
PFT	109	0.65	0.18	0.20–1.00	−0.36	−0.32
MRT	109	0.54	0.19	0.12–0.96	−0.01	−0.37
CC	109	0.67	0.18	0.17–1.00	−0.56	−0.20
CT (ACC)	86	0.90	0.08	0.71–1.00	−0.74	−0.35
CT (RT)	86	2,993	1,153	1,427–8,507	2.01	6.15
R-Cube-SR-Plain (ACC)	109	0.94	0.07	0.67–1.00	−1.61	3.21
R-Cube-SR-Plain (RT)	109	4,494	1,337	2,446–9,099	1.04	1.17
R-Cube-SR-Pattern (ACC)	104	0.81	0.12	0.45–1.00	−0.60	−0.15
R-Cube-SR-Pattern (RT)	78	7,022	2,362	2,392–15,822	1.08	1.71
R-Cube-Vis (wACC)	101	2.77	0.53	1.67–3.93	−0.16	−0.58

Note. ACC = accuracy over all items; CC = Cube Comparison test; CT = Chronometric test; *M* = mean; MRT = Mental Rotation Test; PFT = Paper Folding Test; RT = reaction time; *SD* = standard deviation; wACC = weighted ACC. The different sample sizes result from different excluded participants for each test.

because the IST-WA showed high correlations with the non-spatial relations tests, PFT and MRT, in Study 1. The R-Cube-Vis Test was used in a short version (Fehring, 2020).

## Method

### Participants

The sample consisted of *N* = 109 participants, 92 female and 17 males. The participants were on average 22 years (*M* = 22.17, *SD* = 2.75), whereby the youngest was 18 and the oldest 34 years. All participants were from a German University and could choose between course credit and monetary compensation. Ten participants had an error rate of more than 30% in the target test (R-Cube-SR, Plain) and were excluded for the following analyses.

### Materials

The PFT, the MRT, the CT, and the R-Cube-SR Test (Plain and Pattern) were administered as in Study 1 (see Materials section). Instead of the IST-WA, the Cube Comparison Test (CC, Figure 1) was administered. According to Ekstrom et al. (1976b), participants had 3 min for each of the two pages of the CC. Each page shows 21 items. One item could be interpreted ambiguously as same or different. Therefore, it was excluded from the following analyses. The total number of items was 41. Participants got one point for each correctly answered item. In contrast to Study 1, the R-Cube-Vis Test was administered as a short version with ten items within each difficulty level (five possible and five impossible), 60 items together. The short version, of which the procedure was adapted in a later study (Fehring, 2020), used two sets of items that were balanced between the participants.

### Procedure

The participants conducted the three paper-pencil tests, PFT, MRT, and CC, followed by the CT. After that, they

**Table 5.** Correlations between the validation tests (Study 2)

	R-Cube-Vis	PFT	MRT	CC
PFT	<b>.45***</b>			
MRT	.28**	<b>.49***</b>		
CC	.23*	<b>.46***</b>	<b>.49***</b>	
CT (RT)	<b>−.40***</b>	<b>−.50***</b>	<b>−.53***</b>	<b>−.55***</b>

Note. CC = Cube Comparison test; CT = Chronometric Test; MRT = Mental Rotation Test; PFT = Paper Folding Test; RT = reaction time. All correlations  $|r| \geq .3$  are bold. \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

performed one of the R-Cube-Vis Test versions. Finally, the participants conducted the R-Cube-SR in its Plain and Pattern-version and answered the descriptive data questionnaire.

## Results

The descriptive results were comparable to the results found in Study 1 (Table 4). The larger number of excluded participants for the CT and the reaction times of the Pattern-version of the R-Cube-SR Test resulted from the larger number of participants with more than 30% error rate. A large error rate was also found for the CC. This might be caused by the more complex items utilized in this test.

At first, only the correlations between the validation tests are considered. The correlation between all validation tests (except for the R-Cube-Vis Test) were equally high and ranged between  $.46 \leq |r| \leq .55$ ,  $p < .001$  (Table 5). The differences between the correlations were not significant,  $|z| \leq 0.83$ ,  $p \geq .203$ . The correlations with the R-Cube-Vis Test decreased as expected from the PFT over the MRT to the CC with a significant difference between the correlations with the PFT and the CC,  $z = 1.72$ ,  $p = .043$ . The correlations with the CT were unexpectedly higher ( $r = -.40$ ,  $p < .001$ ), but was still descriptively lower than the correlations between the CT and the other not-spatial relations tests, PFT and MRT (Table 5). The multiple imputation resulted in the same correlations ( $|\Delta r| \leq .02$ ). Four of the



**Table 6.** Correlations of the R-Cube-SR Test versions with the validation tests (Study 2)

Measure	R-Cube-Vis	PFT	MRT	CC	CT
R-Cube-SR-Plain					
RT	-.08	-.31**	-.47***	-.51***	.55***
R-Cube-SR-Pattern					
ACC	.32**	.37***	.34***	.36***	-.27*
RT	.26*	-.13	-.34**	-.23*	.44***

Note. ACC = accuracy over all items; CC = Cube Comparison test; CT = Chronometric test; MRT = Mental Rotation Test; PFT = Paper Folding Test; RT = reaction time. All correlations  $|r| \geq .3$  are bold. \* $p < .05$ ; \*\*\* $p < .001$ .

five correlations between the validation tests and the grade in Mathematics were significant,  $r = -.24$ ,  $p = .016$  (R-Cube-Vis);  $r = -.43$ ,  $p < .001$  (PFT);  $r = -.25$ ,  $p = .008$  (CC);  $r = .34$ ,  $p < .001$  (CT);  $r = -.17$ ,  $p = .074$  (MRT). All correlations between the validation tests and the grade in German as well as experience with Rubik's cubes were not significant,  $|r| \leq .11$ ,  $p \geq .266$  with a marginal significant correlation between the R-Cube-Vis Test and the grade in German,  $r = .19$ ,  $p = .058$ .

The reliability estimates of the reaction times of the R-Cube-SR Test were good,  $\omega = .87$  [.83; .90] (Plain),  $\omega = .90$  [.87; .94] (Pattern). The accuracy measure of the R-Cube-SR Test (Pattern) was poor with  $\omega = .62$  [.46; .79].

The reaction times of the R-Cube-SR Test (Plain) showed the expected correlations with the validation tests (Table 6), that is, equally high correlations with both standard tests for spatial relations, CC and CT,  $.51 \leq |r| \leq .55$ ,  $p < .001$ , and significantly lower correlations with the visualization tests, PFT ( $z \geq 1.81$ ,  $p = .035$ ) and R-Cube-Vis ( $z \geq 3.44$ ,  $p < .001$ ). The correlation with the MRT was descriptively between these correlations, but significant differences were not obtained ( $z \leq 0.74$ ,  $p \geq .229$ ).

The reaction times of the R-Cube-SR Test (Pattern) also showed the expected pattern with a higher correlation with the CT (spatial relations test), a lower correlation with the MRT, and nearly no correlation with the PFT (visualization). The correlation with CT was also significantly larger than the correlation with the PFT,  $z = 1.98$ ,  $p = .024$ . However, the correlations with R-Cube-Vis and CC were comparable. All correlations were descriptively lower than the Plain-version (except for the R-Cube-Vis Test), and there was a marginal significant difference between the correlations with both spatial relations tests (CC and CT),  $z = 1.41$ ,  $p = .079$ . All correlations between the accuracy measure of the R-Cube-SR Test (Pattern) and the validation tests were in a lower range,  $.27 \leq |r| \leq .37$ ,  $p < .016$ . All correlations resulting from the multiple imputation were comparable ( $|\Delta r| \leq .03$ ), except for the correlation between the reaction times of the Pattern version with the R-Cube-Vis Test ( $r = .16$ ,  $p = .230$ ) and with the CT ( $r = .38$ ,  $p = .002$ ). As in Study 1, the correlations were also computed for the only-female

subsample. Also, here, all correlations differed with  $|\Delta r| \leq .10$ , except for the correlations between the reaction times of R-Cube-SR Test (Pattern) and PFT ( $r = .04$ ,  $p = .769$ ). The overall correlation pattern was the same.

The confirmatory factor analysis (CFA) that assigned the R-Cube-SR Test (Plain) to the latent Spatial Relations variable (Figure 6) had the best fit to the data compared to the other two CFAs (see Figure 4);  $\text{AIC}[\text{CFA } 3] = -101.54 < \text{AIC}[\text{CFA } 2] = -99.74 < \text{AIC}[\text{CFA } 1] = -99.36$ ;  $\text{BIC}[\text{CFA } 3] = -68.02 < \text{BIC}[\text{CFA } 1] = -65.84 < \text{BIC}[\text{CFA } 2] = -63.82$ . The fit indices showed acceptable (RMSEA = .07, [.00; .16]) to good fit (CFI = .98; TLI = .96). In contrast, the CFA assigning the R-Cube-SR Test to the middle of the continuum (latent variable MiddleVisSR, see Figure 4) was the best for the Pattern-version according to AIC, but not to BIS ( $\text{AIC}[\text{CFA } 2] = -37.66 < \text{AIC}[\text{CFA } 3] = -35.72 < \text{AIC}[\text{CFA } 1] = -33.92$ ;  $\text{BIC}[\text{CFA } 3] = -2.91 < \text{BIC}[\text{CFA } 2] = -2.50 < \text{BIC}[\text{CFA } 1] = -1.11$ ). The fit indices were acceptable to poor, RMSEA = .16 [.09; .24], CFI = .90, TLI = .74.

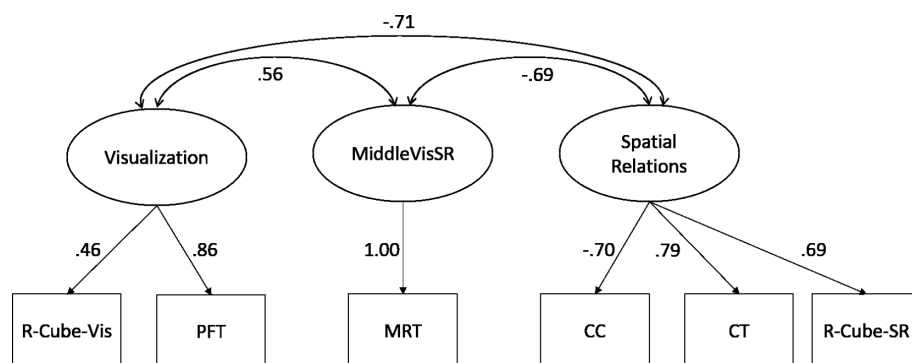
Only the reaction times of the R-Cube-SR Test (Plain) and the accuracy of the R-Cube-SR Test (Pattern) were significantly correlated with the grade in Mathematics,  $r = .21$ ,  $p = .030$  (Plain, RT);  $r = -.34$ ,  $p < .001$  (Pattern, ACC);  $r = -.13$ ,  $p = .264$  (Pattern, RT). All correlations with the grade in German and experience with Rubik's cubes were not significant ( $|r| \leq .14$ ,  $p \geq .141$ ).

## General Discussion

The goal of the present studies was to develop and validate a new spatial thinking test measuring spatial relations that can be distinguished from the spatial thinking factor visualizations.

Existing tests of spatial relations reach similar high correlations with visualization tests compared to correlations with each other. One reason might be the complexity of the visual items of these tests that might involve visualization ability. The presented R-Cube-SR Test in its Plain-version consists of very simple items (solved Rubik's cubes) which have to be rotated as a whole. A more complex version (Pattern), which is more similar to the standard spatial relations tests, was also developed and tested as an opportunity to compare simple and complex variants of a spatial relations test based on the same stimulus materials.

Two studies were conducted to analyze the correlation patterns with standard spatial thinking tests (validation tests) with respect to discriminant and convergent evidence of validity as well as correlations with grades in Mathematics and German as validity evidence for external criteria. The validation tests were the Paper Folding Test (PFT,



**Figure 6.** Confirmatory factor analysis with the R-Cube-SR Test (Plain) assigned to the latent Spatial Relations variable (Study 2).

Ekstrom et al., 1976a) and the new developed R-Cube-Vis Test for visualization and the Chronometric Test (CT, Jansen-Osmann & Heil, 2007), the Würfelaufgaben of the Intelligenz-Struktur-Test (IST-WA, Amthauer et al., 1999, Study 1) and the Cube Comparison Test (CC, Ekstrom et al., 1976a, Study 2) for spatial relations. Furthermore, the correlations with the Mental Rotation Test (MRT, Vandenberg & Kuse, 1978) as a test in the middle of the continuum from spatial relations to visualization were also analyzed.

The correlation patterns of the validation tests in both studies confirmed the results reported in the literature (e.g., Blajenkova et al., 2006; Kozhevnikov et al., 2002; Kozhevnikov & Hegarty, 2001). There were high correlations between tests of the same spatial thinking factor (visualization, spatial relations), but also comparable correlations between tests from the two different factors. The R-Cube-Vis Test measuring visualization was the only test that showed a differentiated correlation pattern in both studies with (marginal) significant lower correlations with spatial relations tests than with the PFT. Only in Study 1, the CT as a spatial relations test showed a significantly higher correlation with the second spatial relations test (IST-WA) than the correlations with the R-Cube-Vis Test.

In contrast, the target test, the R-Cube-SR Test (Plain), with reaction times as the ability measure showed high convergent validity with the Chronometric Test (Studies 1 and 2) and the Cube Comparison test (Study 2) as well as significantly lower correlations with the visualization tests, the R-Cube-Vis Test and the PFT (only in Study 2) as evidence for discriminant validity. As expected, the correlations with the MRT were descriptively lower than the correlations with the spatial relations tests, but higher than the correlations with the visualization test. The Pattern-version of the R-Cube-SR Test with reaction times as the ability measure showed the expected correlation pattern with the validation tests but seemed less suitable to measure spatial relations than the Plain-version (e.g., there were descriptively different correlations between both spatial relations tests, CT and CC, in Study 2, and a considerable number

of participants had to be excluded with error rates exceeding 30%). The Pattern-version with accuracy as the ability measure showed an indifferent correlation pattern with a tendency to higher correlations with the visualization tests, especially in Study 2. Hence, the reaction times of the Plain-version of the R-Cube-SR Test were identified as the most promising measure (of all three tested measures) for spatial relations. The R-Cube-SR Test of both versions with reaction times as the measure had good reliability estimates in both studies.

The correlations patterns in both studies were supported by the conducted confirmatory factor analyses (CFAs). The analyses showed that the CFA assigning the R-Cube-SR Test (Plain) to the latent Spatial Relations variable had better fit the data compared to both alternative CFAs that assigned the R-Cube-SR Test (Plain) to the Visualization end or in the middle of continuum. These results show that the R-Cube-SR Test (Plain) can be interpreted as a test for spatial relations and not for visualization.

The correlations with the external criteria, the grades in Mathematics and German, showed the expected relation with the grade in Mathematics (Gunderson et al., 2012; Hegarty & Kozhevnikov, 1999) in Study 2. The non-significant correlations in Study 1 with both grades as well as with the grade in German in Study 2 do not weaken the evidence of the validity of external criteria, because the same correlation pattern was also found for the standard tests for spatial relations, CT, IST-WA, and CC.

One limitation of the present study is that only two external criteria were considered in present study. Further criteria should be addressed in the following studies regarding predictive power. Another limitation of the present study is the specific spatial thinking tests chosen for convergent and discriminant validation. However, these tests are the current standard tests for the considered spatial thinking factors, and they are commonly used in research and diagnosis of spatial thinking. Moreover, the literature provided results on correlation patterns to derive expectations for the present study. Another potential limitation of the R-Cube-SR Test is that the colored cubes are not suitable

for color-blind individuals. However, the proportion of color-blind individuals is around 5%, which means that the test can be applied for around 95% of the population. A further limitation is the uneven proportion of female and male participants in the samples. However, the single correlations as well as the correlation patterns in both studies were analyzed for the female subsample and the results were the same.

The present study provides evidence for convergent and discriminant validity for the Plain-version of the R-Cube-SR Test (reaction times) and demonstrates the advantage of this test compared to the standard tests of spatial relations by its clear correlation pattern. In particular, the correlations with the R-Cube-Vis Test for visualization that uses similar visual figures as the stimulus materials were low. Furthermore, valid evidence for external criteria was found for the grade in Mathematics. To this end, the Plain-version of the R-Cube-SR Test with reaction times as the ability measure is a valid test for spatial relations and overcomes the disadvantages of standard spatial relations tests by its ability to distinguish between spatial relations and visualization. The R-Cube-Vis Test (Fehringer, 2020) and the R-Cube-SR Test together provide a test set that is able to deliver a differentiated profile of a participants' spatial thinking abilities.

## References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (1999). *Intelligenz-Struktur-Test 2000 (I-S-T 2000)* [Intelligence Structure Test]. Hogrefe.
- Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new Self-Report Imagery Questionnaire. *Applied Cognitive Psychology*, 20(2), 239–263. <https://doi.org/10.1002/acp.1182>
- Burton, L. J., & Fogarty, G. J. (2003). The factor structure of visual imagery and spatial abilities. *Intelligence*, 31(3), 289–318. [https://doi.org/10.1016/S0160-2896\(02\)00139-3](https://doi.org/10.1016/S0160-2896(02)00139-3)
- Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS One*, 10(3), Article e0121945. <https://doi.org/10.1371/journal.pone.0121945>
- Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976a). *Kit of Factor-Referenced Cognitive Tests*. Educational Testing Service.
- Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976b). *Manual for Kit of Factor-Referenced Cognitive Tests*. Educational Testing Service.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fehringer, B. C. O. F. (2020). Spatial thinking from a different view: Disentangling top-down and bottom-up processes using eye tracking. *Open Psychology*, 2, 138–212. <https://doi.org/10.1515/psych-2020-0105>
- Fehringer, B. C. O. F. (2021a). *Implementation of the R-Cube-SR Test in English as well as German*. PsychArchives. <https://doi.org/10.23668/PSYCHARCHIVES.5160>
- Fehringer, B. C. O. F. (2021b). *Supplementary material to "R-Cube-SR Test: A new test for spatial relations distinguishable from visualization"*. <https://doi.org/10.23668/psycharchives.4964>
- Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology*, 48(5), 1229–1241. <https://doi.org/10.1037/a0027433>
- Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology*, 91(4), 684–689. <https://doi.org/10.1037/0022-0663.91.4.684>
- Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, 32(2), 175–191. <https://doi.org/10.1016/j.intell.2003.12.001>
- Hegarty, M., & Waller, D. A. (2005). Individual differences in spatial abilities. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 121–169). Cambridge University Press.
- Henn, P., Gallagher, A. G., Nugent, E., Seymour, N. E., Haluck, R. S., Hseino, H., Traynor, O., & Neary, P. C. (2018). Visual spatial ability for surgical trainees: Implications for learning endoscopic, laparoscopic surgery and other image-guided procedures. *Surgical Endoscopy*, 32(8), 3634–3639. <https://doi.org/10.1007/s00464-018-6094-3>
- Jansen, P., Schmelter, A., Quaiser-Pohl, C., Neuburger, S., & Heil, M. (2013). Mental rotation performance in primary school age children: Are there gender differences in chronometric tests? *Cognitive Development*, 28(1), 51–62. <https://doi.org/10.1016/j.cogdev.2012.08.005>
- Jansen-Osmann, P., & Heil, M. (2007). Suitable stimuli to obtain (no) gender differences in the speed of cognitive processes involved in mental rotation. *Brain and Cognition*, 64(3), 217–227. <https://doi.org/10.1016/j.bandc.2007.03.002>
- Kelley, K. (2020). *MBESS: The MBESS R package* (R package version 4.8.0). <https://CRAN.R-project.org/package=MBESS>
- Kline R. B. (Ed.). (2016). *Methodology in the social sciences. Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition*, 29(5), 745–756. <https://doi.org/10.3758/BF03200477>
- Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction*, 20(1), 47–77. [https://doi.org/10.1207/S1532690XCI2001\\_3](https://doi.org/10.1207/S1532690XCI2001_3)
- Kozhevnikov, M., Kozhevnikov, M., Yu, C. J., & Blazhenkova, O. (2013). Creativity, visualization abilities, and visual cognitive style. *The British Journal of Educational Psychology*, 83(Pt 2), 196–209. <https://doi.org/10.1111/bjep.12013>
- McDonald, R. (1999). *Test theory: A unified treatment*. Erlbaum.
- McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, 86(5), 889–918. <https://doi.org/10.1037/0033-2909.86.5.889>

- Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, 130(4), 621–640. <https://doi.org/10.1037/0096-3445.130.4.621>
- Münzer, S. (2012). Facilitating spatial perspective taking through animation: Evidence from an aptitude–treatment–interaction. *Learning and Individual Differences*, 22(4), 505–510. <https://doi.org/10.1016/j.lindif.2012.03.002>
- Münzer, S., Fehring, B. C. O. F., & Köhl, T. (2018). Specificity of mental transformations involved in understanding spatial structures. *Learning and Individual Differences*, 61, 40–50. <https://doi.org/10.1016/j.lindif.2017.11.004>
- Paivio, A. (2014). Intelligence, dual coding theory, and the brain. *Intelligence*, 47, 141–158. <https://doi.org/10.1016/j.intell.2014.09.002>
- Pässler, K., Beinicke, A., & Hell, B. (2015). Interests and intelligence: A meta-analysis. *Intelligence*, 50, 30–51. <https://doi.org/10.1016/j.intell.2015.02.001>
- Pellegrino, J. W., Alderton, D. L., & Shute, V. J. (1984). Understanding spatial ability. *Educational Psychologist*, 19(3), 239–253. <https://doi.org/10.1080/00461528409529300>
- Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse Mental Rotations Test: Different versions and factors that affect performance. *Brain and Cognition*, 28(1), 39–58. <https://doi.org/10.1006/brcg.1995.1032>
- R Core Team. (2017). *R: A language and environment for statistical computing*. <https://www.R-project.org/>
- Revelle, W. (2017). *Psych: Procedures for psychological, psychometric, and personality research*. <http://CRAN.R-project.org/package=psych>
- Robitzsch, A., & Grund, S. (2020). *miceadds: Some additional multiple imputation functions, especially for R package version 3.9-14*. <https://CRAN.R-project.org/package=miceadds>
- Rossee, Y. (2012). Lavan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Thurstone, T. G. (1958). *Manual for the SRA primary mental abilities*. Science Research Associates.
- Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology*, 7, Article 769. <https://doi.org/10.3389/fpsyg.2016.00769>
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multi-variate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://www.jstatsoft.org/v45/i03/>
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47(2), 599–604. <https://doi.org/10.2466/pms.1978.47.2.599>
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for stem domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817–835. <https://doi.org/10.1037/a0016127>
- Wang, L., & Carr, M. (2014). Working memory and strategy use contribute to gender differences in spatial ability. *Educational Psychologist*, 49(4), 261–282. <https://doi.org/10.1080/00461520.2014.960568>
- Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12), 1–20. <http://www.jstatsoft.org/v21/i12/>
- Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1), 1–29. <http://www.jstatsoft.org/v40/i01/>
- Wickham, H., François, R., Henry, L., & Müller, K. (2018). *dplyr: A grammar of data manipulation* (R package version 0.7.6). <https://CRAN.R-project.org/package=dplyr>

## History

Received July 24, 2021

Revision received July 9, 2021

Accepted August 13, 2021

Published online November 10, 2021

EJPA Section/Category Intelligence

## Acknowledgment

I am very thankful for the many, very supportive discussions about the manuscript with Stefan Münzer, who provided valuable feedback to this work.

## Open Science

Open Data: The information needed to reproduce all of the reported results are not openly accessible.

Open Materials: The information needed to reproduce all of the reported methodology is not openly accessible. The supplementary material is available at <http://dx.doi.org/10.23668/psycharchives.4964> (Fehring, 2021b). An implementation of the R-Cube-SR Test is available at <https://doi.org/10.23668/PSYCHARCHIVES.5160> (Fehring, 2021a).

Preregistration: This study was not preregistered.

## Funding

This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences funded by the German Research Foundation. Open access publication enabled by the University of Mannheim.

## Benedict C. O. F. Fehring

Department of Psychology of Education

University of Mannheim

A5 6

68159 Mannheim

Germany

[b.fehring@uni-mannheim.de](mailto:b.fehring@uni-mannheim.de)