Rotational complexity in mental rotation tests

Nolte, Nils; Schmitz, Florian; Fleischer, Jens; Bungart, Maximilian; Leutner, Detlev

Published in:
Intelligence

DOI:
10.1016/j.intell.2022.101626

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Intelligence 91 (2022) 101626

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Rotational complexity in mental rotation tests: Cognitive processes in tasks requiring mental rotation around cardinal and skewed rotation axes

Nils Nolte[a,*], Florian Schmitz[a], Jens Fleischer[b], Maximilian Bungart[c], Detlev Leutner[a]

[a] Institute of Psychology, University of Duisburg-Essen, Universitätsstr. 2, 45141 Essen, Germany
[b] Institute of Educational Science, Ruhr-University Bochum, Universitätsstraße 150, 44801 Bochum, Germany
[c] Institute of Psychology, Saarland University, Campus, 66123 Saarbrücken, Germany

ARTICLE INFO

Keywords:
Mental rotation
Spatial ability
Vandenberg & Kuse paradigm
Rotation axes
Cognitive processes

ABSTRACT

Mental rotation tests have been extensively studied regarding item characteristics that affect difficulty, e.g., angular disparity, item dimensionality, and object complexity. In the present study, using the Vandenberg and Kuse (1978) paradigm, we applied a psychometric approach to examine whether complex skewed-axis rotation requires incremental processes that can be distinguished from simple cardinal-axis rotation. Participants (N = 372) completed a battery of cognitive tests, including a mental-rotation test requiring mental rotation of Shepard and Metzler type figures around a single cardinal axis or around two cardinal axes (resulting in a skewed-axis rotation). When comparing a nested-factor measurement model to a one-factor model, results showed that complex skewed-axis rotation is not identifiable as a nested specific factor. This suggests that the processes resulting in individual differences in mental rotation are either the same in both item types, or at least substantially correlated. Including spatial visualization tests and reasoning tests in a prediction model suggested that participants used spatial strategies over and above reasoning to solve the mental rotation items. These results generalize the findings of Just and Carpenter (1985) on simple cardinal-axis and complex skewed-axis rotation of cubes to more complex objects that allow more flexible mental rotations. It can be concluded that mental rotation represents a unitary ability. From an individual-differences perspective, this ability can be assessed equally with simple cardinal-axis and complex skewed-axis rotation items.

1. Introduction

Mental rotation tasks have received considerable attention in ability research. As early as 1938, Thurstone conceived of spatial ability as a primary mental ability, and a corresponding factor is also included in Carrol's model (Carroll, 1993) as well as in the most recent taxonomy of human abilities, the CHC model (Flanagan & Dixon, 2013). Further, mental rotation tests have been used as ability indicators because they are closely related with reasoning, fluid intelligence (Carroll, 1993), and the general factor g of intelligence (Lohman, 1996), although they appear to be dissociable from the latter (Carroll, 1993; Flanagan & Dixon, 2013; Thurstone, 1938). According to Carroll (1993), mental rotation is part of the spatial visualization facet of spatial ability, which he defined as “processes of apprehending, encoding, and mentally manipulating spatial forms”.

Experimental cognitive psychology has shed light on low-level processes that are assumed to affect performance in mental rotation tests. These processes relate, among others, to task characteristics such as angular disparity (e.g., R. N. Shepard & Metzler, 1971, for 3D tasks, and Cooper, 1975, for 2D tasks), stimulus dimensionality (S. Shepard & Metzler, 1988), or axis of rotation (Parsons, 1987). It is less known, however, to what extent these task characteristics are relevant to the assessment of individual differences in mental rotation ability, since they may call for different cognitive functions and abilities. If this were the case, it could affect how mental rotation should be assessed for diagnostic purposes.

The aim of the present contribution is to investigate mental rotation ability as assessed with the Vandenberg and Kuse paradigm (S. G. Vandenberg & Kuse, 1978) which is one of the most frequently used rotation tests in applied assessment. In this introduction we start by reviewing the relevance of the paradigm in research and its validity for ability assessment. Next, adopting a cognitive processing perspective we discuss task requirements and processing strategies; a particular focus is on the complexity of mental rotation as determined by the axes of

* Corresponding author.
E-mail address: nils.nolte@uni-due.de (N. Nolte).

https://doi.org/10.1016/j.intell.2022.101626
Received 30 April 2021; Received in revised form 8 November 2021; Accepted 5 January 2022
Available online 5 February 2022
0160-2896/

rotation. We then summarize the evidence available today concerning individual differences. Finally, we elaborate on the aim of this contribution, namely: to explore (1) the dimensionality of mental rotation ability from an assessment perspective and (2) its incremental validity over and above reasoning ability.

1.1. The Vandenberg and Kuse paradigm

The Vandenberg and Kuse paradigm (S. G. Vandenberg & Kuse, 1978) is a highly popular measure of mental rotation ability. In this task, the classic Shepard and Metzler figures (R. N. Shepard & Metzler, 1971) are used. These are 3-dimensional figures composed of conjoined regular cubes (see Fig. 1). Participants are instructed to inspect an anchor stimulus and to select two identical but rotated target stimuli from a row of stimuli including two distractors (which are either mirrored or differently composed). Performance is measured as the percentage of correct answers ("found both target stimuli in the item") across a series of items, and usually scores around 40–50% accuracy (e.g., Krüger & Suchan, 2016; Peters, 2005; S. G. Vandenberg & Kuse, 1978). This is considerably less than the 95% accuracy level frequently obtained in the Shepard and Metzler paradigm (R. N. Shepard & Metzler, 1971). Consequently, the Vandenberg and Kuse paradigm does not suffer from ceiling effects; this is beneficial for the assessment of individual differences.

In fact, the Vandenberg and Kuse paradigm has been frequently used in studies on spatial ability and its relevance in applied contexts: For instance, it has been shown to be a predictor of academic and professional success in science, technology, engineering, mathematics (i.e., the STEM fields), and medicine, ranging from study achievement (e.g., Sorby, Veurink, & Streiner, 2018) to dentistry (Hegarty, Keehner, Khooshabeh, & Montello, 2009), endoscopy (Judd & Klingberg, 2021), or aviation security (Krüger & Suchan, 2016). Furthermore, the paradigm has been used in studies on sex differences in rotation ability, which revealed that male participants typically outperform female participants (for a review see Voyer, Voyer, & Bryden, 1995). Hormones have been discussed as a possible cause of the observed group differences (Hausmann et al., 2000).

The composition of the test sets varies between studies, especially since Peters and Battista's (2008) stimulus library has enabled researchers to easily create their own test sets.

1.2. Task characteristics

Performance in ability tests generally depends on task characteristics. For mental rotation tests, some of the most important of these are, among others, angular disparity, stimulus dimensionality and axis of rotation.

Probably the best-documented factor is angular disparity. R. N. Shepard and Metzler (1971) first described a consistent linear relationship between angular disparity and reaction time, which other researchers confirmed under different circumstances (e.g., Cooper, 1975, for 2D tasks). The common explanation for this is the overlap between mental rotation and manual rotation found in fMRI studies (for a review, see Zacks, 2008). Boone and Hegarty (2017) described a trend similar to that for reaction time for accuracy up to 90° of angular disparity.

In contrast, the rotated stimulus' dimensionality appears to mainly affect the intercept of the reaction-time-by-angular-disparity function (S. Shepard & Metzler, 1988), indicating that dimensionality predominantly affects the encoding of the stimulus and, to a lesser extent, the actual rotation process.

The required axis of rotation affects the slope of the reaction-time-by-angular-disparity function, with the horizontal axis displaying the shallowest slope and the line-of-sight axis, perpendicular to the picture plane, displaying the steepest slope (Parsons, 1987). However, Stieff et al. (2018) showed that this order changes when accuracy is considered, instead of reaction time: Rotation in the picture plane – that is, a rotation around an axis perpendicular to the picture plane – was more accurate than any rotation in depth, that is, a rotation around an axis not perpendicular to the picture plane.

Other task characteristics of interest include stimulus complexity (e.g., Cooper, 1975; Heil & Jansen-Osmann, 2008; Stieff, 2007; Stieff et al., 2018) and the participant's familiarity with the type of stimulus (e.g., Doyle & Voyer, 2018; Muto & Nagai, 2020; Stieff, 2007).

Another characteristic affecting performance that is not intrinsic to the stimuli, is the application of a time limit, as is commonly applied in the Vandenberg and Kuse paradigm: Shorter time limits lead to weaker performance (Peters, 2005). Time limits also exacerbate gender differences: In a meta-analysis, Voyer (2011) found a linear relationship between the length of the time limit and the magnitude of sex differences, and the smallest sex difference in studies with no time limit. The effects of time limits may be mediated by strategy choice, with participants adopting more time-efficient strategies being less impacted by a time limit.

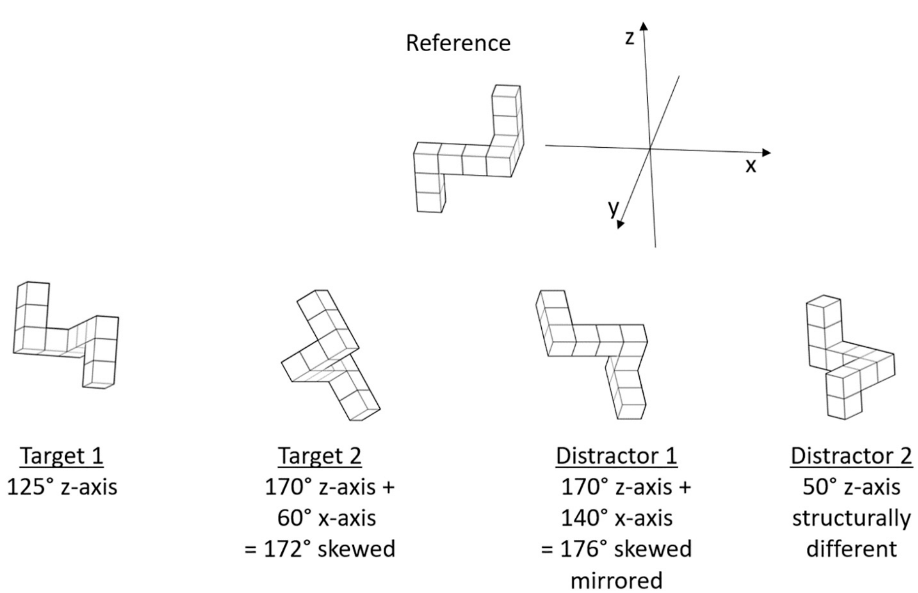

Fig. 1. Example item of the Vandenberg & Kuse mental rotation test showcasing the reference frame spanned by the figure segments and displaying the response alternatives.
Note. Target 1 represents a simple cardinal-axis rotation in relation to the reference figure while Target 2 represents a complex skewed-axis rotation in relation to the reference figure. Distractor 1 is a mirror foil while Distractor 2 is a structure foil.

1.3. Processing strategies

From a cognitive processing perspective, a core question pertains to what strategies are applied when participants complete a test. For mental rotation, two strategies have been proposed as most prevalent, namely holistic (spatial) and analytical (reasoning-based; Burin, Delgado, & Prieto, 2000).

The holistic spatial strategy is expected for mental rotation tasks as it necessitates the imagination of rotating the shown stimulus (i.e., mentally rotating it; Burin et al., 2000). When this strategy is used, the test is expected to measure spatial visualization defined as "processes of apprehending, encoding, and mentally manipulating spatial forms" (Carroll, 1993).

Opposed to that, analytical strategies in spatial tasks are mainly centered around feature matching or feature comparison (e.g., Glück, Machat, Jirasko, & Rollett, 2002). This is an analytical approach that requires reasoning instead of spatial ability (in contrast to the holistic spatial approach that requires mental rotation). Participants using an analytical strategy focus on salient features to aid their same-different decision – most frequently the shape of either the whole stimulus or separate parts of the stimulus is used (Burin et al., 2000), but other features such as color patterns are also possible (Khooshabeh & Hegarty, 2010).

It has been argued that in the Vandenberg and Kuse paradigm participants often use both types of strategies combined (Hegarty, 2018). This may be prompted by items that have two types of distractors: on the one hand, mirrored versions of the target stimulus (mirror foils), and on the other hand stimuli differing slightly in shape (structure foils). Structure foils can be eliminated using feature matching, while mirror foils cannot. Therefore, participants predominantly using analytic strategies such as feature matching perform much worse on items having only mirror foils than on items having only structure foils (Geiser, Lehmann, & Eid, 2006).

Time limits interact with strategy choice, as holistic strategies are usually more time efficient and therefore lead to increased performance under restricted time, compared to analytical strategies. This is another possible reason for the male advantage in the Vandenberg and Kuse paradigm, which is usually applied with a time limit: Compared to women, men are more likely to use efficient strategies (Hirnstein, Bayer, & Hausmann, 2009) and are less likely to use more analytical strategies (such as feature matching; Geiser et al., 2006).

Stimulus complexity and axes of rotation also affect strategy choice. While a 90° rotation around one edge of a simple cube may be represented as turning the cube over one step, rotations of more complex stimuli and around less pre-defined rotational axes can lead to different types of rotational strategy (Just & Carpenter, 1985; Parsons, 1987) or, due to increased difficulty, to switching from rotation to feature matching, as Boone and Hegarty's (2017) results on changes in response strategy in items with high angular disparity indicate.

1.4. Cardinal axes

With the use of stimuli composed of regular cubes, cardinal axes corresponding to the edges of the cubes pop out relatively easily. Similarly, in the Shepard and Metzler figures the cardinal axes are given by the spatial orientation of segments of the figures, as the segments are orthogonal to each other and thereby span a 3D reference frame for rotation (see Fig. 1). In the original R. N. Shepard and Metzler (1971) studies and in most similar studies, figures were rotated around one of these cardinal axes.

However, mental rotation in the real world is often considerably more complex, and requires rotation around a skewed axis that does not correspond to any of the cardinal axes defined by the edges of the object. Just and Carpenter (1985) and Parsons (1987) both describe two types of rotational strategy that may be pursued to this end: On the one hand, participants could use a multiple-step rotation ("rotation by dimension") around the cardinal axes. On the other hand, participants may rotate a stimulus in a single-step (or complex skewed) rotation, which is referred to by Parsons as "shortest path rotation". In multiple-step rotation the rotation axes are given by the cardinal axes of the reference frame, whereas in single-step rotation, the skewed rotation axis is not given by and does not coincide with the cardinal axes of the reference frame. In this latter case, the skewed rotation axis has to be identified by the participant, allowing for a direct, single-step rotation.

As could be expected, items requiring rotation around a skewed axis showed longer reaction times compared to items including the same degree of rotation but around a single cardinal axis (Just & Carpenter, 1985; Parsons, 1987). This can be explained in relation to both rotation strategies: Multiple-step "rotation by dimensions" takes longer as it entails a longer rotation path as well as a switch between rotation axes. The direct "shortest path" rotation also takes longer than rotation around a single cardinal axis as it requires identification of the item-specific skewed rotation axis before performing the single-step rotation. The need to first identify a rotation axis is an additional requirement in the mental rotation process that makes such tasks more complex. In line with this, Ziemek, Creem-Regehr, Thompson, and Whitaker (2012) showed that participants performed with higher accuracy on items with mechanical stimuli whose features were aligned to the axes of rotation than items with anatomical stimuli that lacked this property.

Just and Carpenter (1985) used eye-tracking measures to show that, in cubes, only high-spatial participants were able to efficiently utilize the most efficient shortest-path-rotation strategy (Parsons, 1987), while low-spatial participants used the less efficient rotation-by-dimension strategy. This indicates that individual differences in spatial visualization ability influence what type of rotation strategy participants use. Likewise, Arendasy and Sommer (2013) showed that if a task becomes too difficult relative to a participant's ability, participants are more likely to use feature matching strategies instead. Just and Carpenter (1985) described a similar notion.

From an assessment perspective, differences in strategies for solving different types of mental rotation items turn out to be a challenge. It seems that items requiring more complex (skewed or multiple-step) rotations require different cognitive processes – i.e., either direct rotation, including identification of the task-specific rotation axis, or stepwise rotation, including a switch of rotation axes – that may be distinguishable from the cognitive processes that are needed to solve classic cardinal rotation items. This raises the question whether solving complex mental rotation items (requiring rotation around axes that are not given by the frame of reference) necessitates spatial abilities that go beyond the spatial abilities that are needed to solve more simple mental rotation items (requiring rotation around cardinal axes that are given by the frame of reference), and to what extent task performance reflects spatial visualization ability at all (rather than a reasoning-related shortcut process such as feature matching; Hegarty, 2018).

2. Aims of the present study

This study aims to contribute to a better understanding of mental rotation ability, its assessment, and its factorial structure. While classic tests of mental rotation have used simple rotation around cardinal axes (see, e.g., S. Shepard & Metzler, 1988), rotation requirements in the real world usually are more complex and may comprise complex skewed-axis rotation (e.g., for body parts, as described in Wohlschläger & Wohlschläger & Wohlschläger, 1998). In turn, the latter can be achieved by finding the most appropriate direct (skewed) rotation axis or by completing a stepwise sequence of cardinal rotations including the requirement to switch between these. In any case, this provokes the question whether complex skewed-axis rotation necessitates cognitive functions and processes that are psychometrically dissociable from those used for simple rotation around cardinal axes. Put differently, we seek to answer the question whether complex skewed-axis rotation requires incremental abilities that can be distinguished from simple cardinal-axis rotation. Besides its theoretical relevance, this question is also of practical relevance for the specific selection of items in mental rotation ability tests, as Peters and Battista's (2008) stimulus library enables researchers to tailor a mental rotation test to specific needs.

Technically, we seek to address this question by means of fitting a nested-factor CFA to indicators requiring either simple cardinal-axis rotation or complex skewed-axis rotation, with a shared general rotation factor capturing variance in all indicators and a nested specific factor accounting only for the incremental requirements in indicators requiring complex skewed-axis rotation. If the nested factor is identified (substantial and statistically significant), this would be interpreted as evidence that complex skewed-axis rotation captures an ability incremental to that involved in simple cardinal-axis rotation items. In turn, a distinguishable skewed-axis rotation ability could potentially possess differential relations and incremental validity for relevant outcome variables.

Further, we consider previous evidence that mental rotation tasks are generally closely related with reasoning ability, as most ability indicators are heavily intercorrelated (Carroll, 1993), and note that cognitive research has shown that participants may resort to reasoning-related processing strategies (e.g., feature matching) instead of using spatial visualization-related processes, especially with difficult items (Arendasy & Sommer, 2013; Just & Carpenter, 1985). These previous findings raise the question whether the currently available tests used to assess mental rotation, such as tests following the Vandenberg and Kuse paradigm, capture it as an ability that can be distinguished from reasoning ability.

We sought to test this analogously by administering reasoning and visualization tests as relevant predictors. Nested-factor CFA was used to specify a general factor of reasoning ability (where reasoning ability may serve as the closest proxy to general intelligence g; Carroll, 1993) accounting for variance in all indicators. This general reasoning factor would likely account for a substantial portion of variance in visualization tests because of the positive manifold inherent in all ability tests (Carroll, 1993; Spearman, 1904). Importantly, a nested visualization-specific factor was tested that accounts for specific variance in visualization tests only. In turn, this nested specific factor captures portions of variance that are unique to visualization, and thus, independent of the general reasoning factor. In turn, the predictions of the nested factor would indicate a contribution from a specific visualization ability.

These factors of reasoning and visualization, respectively, were tested as predictors of the ability factors modeled for the mental rotation tests. A contribution from the general reasoning factor would be predicted (i.e., reflecting positive manifold; Carroll, 1993; Spearman, 1904). More importantly, we sought to test whether there is visualization-specific ability in those factors accounting for the performance in mental rotation tests. This would indicate that participants indeed use holistic spatial strategies and not only reasoning-based analytical strategies in solving mental rotation tasks – both for stimuli requiring simple cardinal-axis rotations and for stimuli requiring complex skewed-axis rotations.

Hypotheses:

1. The nested-factors measurement model on the criterion side (depicted on the right-hand side in Fig. 2a) fits the data better than the one-factor measurement model (depicted on the right-hand side in Fig. 2b), indicating that including complex skewed-axis rotation as a separate factor provides incremental value for measuring mental rotation ability.

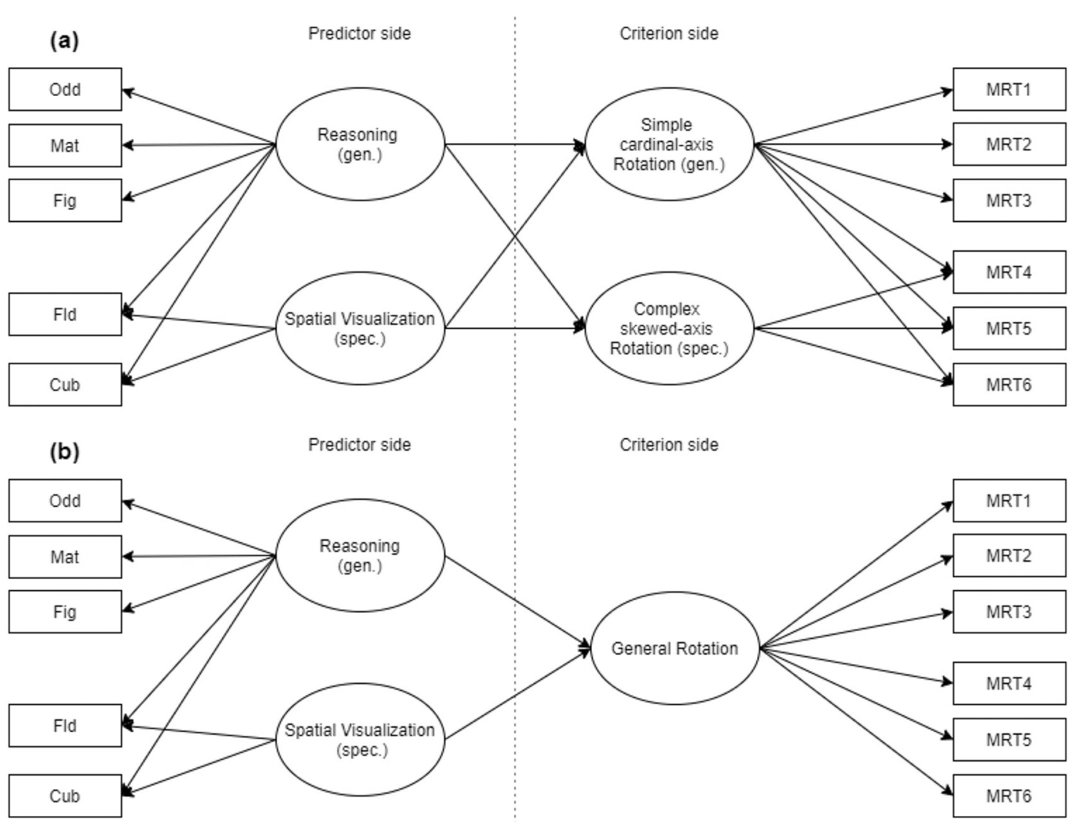

Fig. 2. Prediction models with either two separate factors, a general and a nested specific factor (a), or a single factor (b) on the criterion side.
Note. MRT1-MRT6 are parcels of the mental rotation test, with MRT1-MRT3 consisting of items requiring simple cardinal-axis rotation and MRT4-MRT6 consisting of items requiring complex skewed-axis rotation. Odd = Odd-Man-Out Task; Mat = Figural Matrices Task; Fig = Figure Classification Task; Fld = Paper Folding Task; Cub = Cube Construction Task.

2. Spatial visualization ability predicts mental rotation ability over and above reasoning ability, indicated by significant path coefficients between the nested specific spatial-visualization factor and the factor (s) accounting for variance in the spatial rotation items, let this either be both the simple cardinal-axis rotation factor and the complex skewed-axis rotation factor (in the case where the model depicted in Fig. 2a is adopted) or a general rotation factor (in the case where the model depicted in Fig. 2b is adopted).

3. Method

3.1. Sample

Participants were recruited online via academic teachers and student representatives of psychology and teacher education programs at two German universities, with participants also being encouraged to disseminate the survey in their social environment. Participants needed to be at least 18 years old, not suffering from any psychiatric or neurological disease, and to have normal or corrected eyesight. Of the 557 participants who began the survey, 372 fitted these inclusion criteria and completed at least one of the six tests with a score above guessing probability. Of these, 85% were younger than 30 years (median: 23, range: 18–63) and 58% were female. Most participants reported being university students (74%), employed (25%), or in apprenticeships (4%). Reporting multiple occupations was possible, but only 7% of the participants did report more than one (all double reports were "university student" and "employed"). All participants received a 10 € coupon for completing the survey.

3.2. Procedure and materials

The study protocol was in accordance with the human subjects guidelines of the Declaration of Helsinki. At the time of data collection, institutional review board approvals were neither required nor customary for these types of studies in Germany. Data collection was conducted online using LimeSurvey (LimeSurvey GmbH, 2019) hosted on a university server. For an overview of the tests used, including item numbers and time limits, see Table 1.

3.2.1. Spatial visualization tests

3.2.1.1. Mental rotation task. Mental rotation stimuli were presented in the Vandenberg and Kuse (1978) paradigm, in which participants are asked to pick the two rotated versions of a reference stimulus (see Fig. 1). Stimuli were taken from the stimulus library of Peters and Battista (2008) and were selected following the rules for distinct items described in Krüger and Suchan (2016): Reference stimuli were selected from the stimulus library, going through the different shapes and angular differences included in the library to get distinct items. The first of the two target stimuli was created by rotating the reference stimulus only around the horizontal x axis in half the items and only around the vertical z axis in the other half of the items (simple cardinal-axis rotation; for an example, see Fig. 1). Mean angular disparity was 118° (range: 110°–125°) for these stimuli. The second target stimulus was created by rotating the reference stimulus around both the horizontal x axis and the vertical z axis. This resulted in a skewed one-step rotation (complex skewed-axis rotation; see Target 2 in Fig. 1). Mean angular disparity (for the skewed one-step rotation) in these items was 115° (range: 77°–175°). The first distractor was a mirrored version of the reference stimulus that, in half of the items, was also rotated around both the x axis and the z axis (resulting in a complex skewed-axis rotation). The second distractor was a structurally different stimulus that was rotated around the z axis in half of the items, and around both the x axis and the z axis in the other items.

For each item, the target and distractor stimuli were displayed in randomized order. No time limit was given for attempting these items.

Responses for both target stimuli were treated as different items in the present study. Thereby, each mental rotation item yielded two bits of information, one on correct choice of the target stimulus at a simple cardinal-axis rotation, the other for correct choice of the target stimulus at a complex skewed-axis rotation of the reference stimulus. This allowed computing two scale scores from the 24 items of the mental rotation task: one by calculating the percentage of correctly chosen target stimuli for simple cardinal-axis (1-axis) rotation, the other by calculating the percentage of correctly chosen target stimuli for complex skewed-axis (2-axis) rotation. Means and standard deviations of the two scales were similar ($M_{simple} = 0.76$, $SD_{simple} = 0.16$; $M_{complex} = 0.75$, $SD_{complex} = 0.17$) and both scales showed good reliability (omega total $\omega_{simple} = .88$, $\omega_{complex} = .89$; after McNeish, 2018).

3.2.1.2. Cube construction task. For a three-dimensional test of spatial visualization we used the distractor-based Cube Construction Task created by Bungart (2018) on the basis of the distractor-free Cube Construction Task of Thissen, Koch, Becker, and Spinath (2018). The test consists of 23 items of increasing difficulty, presented successively. In each item, a reference cube is presented from three different perspectives and the respondent needs to identify the correct unfolded cube surface among six options (see Fig. 3). Difficulty is manipulated by varying the number of given symbols on the unfolded sheets. Empty faces on the unfolded cube surfaces could represent any symbol on the reference cube. Item difficulty is greatest with three empty faces and three given symbols.

We administered a shortened version of the original test with 10 items and a time limit of 1 min for each item. Items were selected for high discrimination and short average response times for test efficiency based on data from the original work of Bungart (2018). In our sample, the test had a mean of 0.32 ($SD = 0.22$) and showed good reliability despite the short time limit and high degree of difficulty ($\omega_{cub} = .79$).

3.2.1.3. Paper folding task. We used a Paper Folding Task as a classic measure of spatial visualization (Carroll, 1993). The test was taken from a battery of cognitive tests by Heller and Perleth (2000). The test had 25 items, a mean of 0.55 ($SD = 0.21$), and good reliability ($\omega_{Fld} = .86$).

3.2.2. Reasoning tests

To measure reasoning ability, we administered a group of tests from established test batteries. This included first a Figure Classification Task (Heller & Perleth, 2000; subscale N1) with 15 items, a mean of 0.58 ($SD = 0.19$) and good reliability ($\omega_{Fig} = .89$). Second, we used an "Odd-Man-Out" Task (Weiß, 2006; subscale 2) with 15 items, a mean of 0.75 ($SD =$

Table 1
Overview of tests administered.

Construct	Test name	n items	Time limit	Source
Spatial visualization	Mental Rotation Task	24	None	S. G. Vandenberg and Kuse (1978)
Spatial visualization	Cube Construction Task	10	1 min per item	Bungart (2018)
Spatial visualization	Paper Folding Task	25	9 min	Heller and Perleth (2000)
Reasoning	Figure Classification Task	15	8 min	Heller and Perleth (2000)
Reasoning	Figural Matrices Task	15	4 min	Weiß (2006)
Reasoning	Odd-Man-Out Task	15	5 min	Weiß (2006)

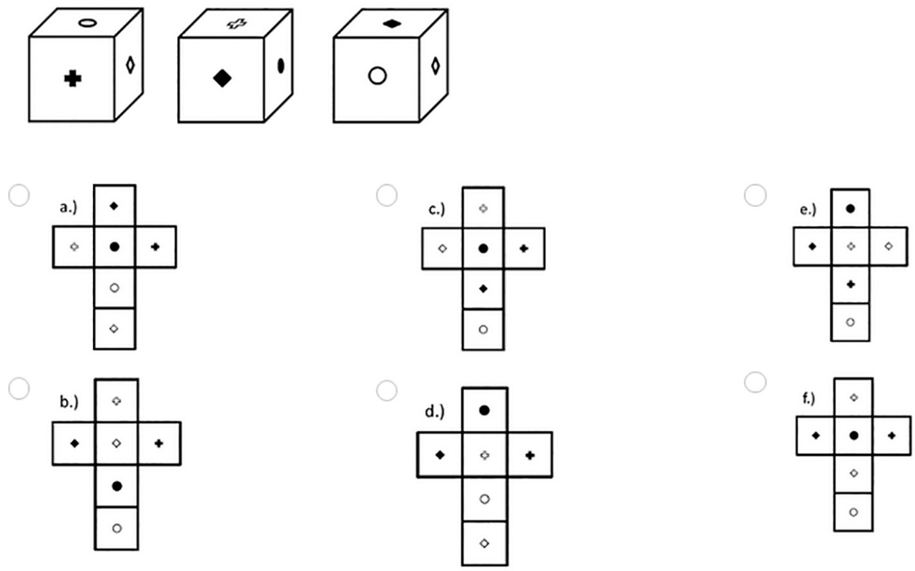

Fig. 3. Example item of the cube construction task with no empty faces on the answer options.

0.15) and $\omega_{Odd} = .68$. Third, we used a Figural Matrices Test from the same battery (Weiß, 2006; subscale 3), similar to Raven's matrices (Raven, 2000). The test had 15 items, a mean of 0.80 (*SD* = 0.16), and good reliability ($\omega_{Mat} = .83$).

3.3. Data preparation

All test items were scored as correct (1) or incorrect (0). If a participant did not complete items, due to skipping single items or to elapsed time limit, these were scored as incorrect if the participant had not skipped the test entirely. If a participant had skipped a test, their score on this test was scored as missing. As we assumed that skipped tests resulted from survey discontinuation unrelated to relevant variables, we treated these missings as "missing at random" and estimated them using a FIML estimator to use all available information from the data (Enders & Bandalos, 2001).

To test a nested-factor measurement model of the mental rotation test (Fig. 4), its items were collapsed to form six parcels: three parcels with simple cardinal-axis rotation items (MRT1-MRT3) and three parcels with complex skewed-axis rotation items (MRT4-MRT6). Each parcel included eight items. Parcels were constructed (100,000 combinations tested) by optimizing measurement invariance in terms of equal loadings, equal intercepts, and equal residual variances of the parcels within the nested-factor measurement models. There were no constraints concerning the overall magnitude of the loadings other than that loadings on the same factor within the measurement model should be equal. The optimization criterion was the least difference between corresponding parameters in terms of a minimum *p* value of the chi-squared difference test. For comparison, all models used in the analyses were also investigated on the basis of parcels optimized only for equality of parameters in their own one-factor models; this yielded highly similar results.

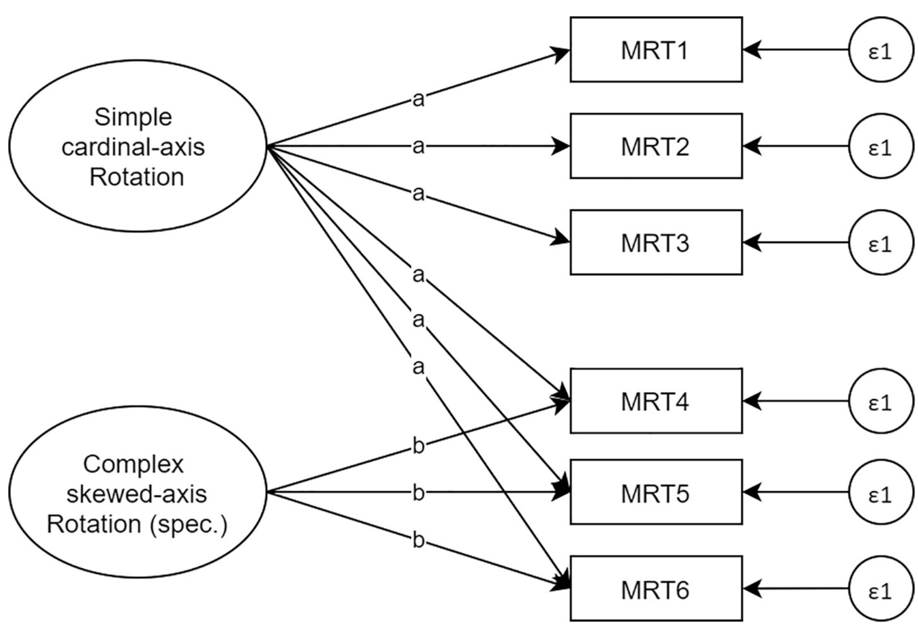

Fig. 4. Measurement model for parcel creation.
Note. MRT1-MRT3 include simple cardinal-axis rotation items, MRT4-MRT6 include complex skewed-axis rotation items.

The proportions of correct responses in each test or test parcel were used as indicators of their respective latent variables. One item had to be excluded for 42 participants, due to programming error; hence, scores for both simple cardinal-axis rotation and complex skewed-axis rotation for these participants were based on the remaining seven items, with a system missing value for the erroneous item.

3.4. Analyses

The difference in item difficulty of the complex skewed-axis and the simple cardinal-axis rotation scales was evaluated via a two-sample one-sided *t*-test (or Welch test, if variance homogeneity was violated).

We used nested-factor measurement models both for predictor variables and criterion variables. Nested-factor models allow for the separation of variance allocated for by a shared general factor and a nested specific factor of interest (see Morin, Myers, & Lee, 2020, for details on bifactor confirmatory factor analysis), leaving only unique variance to the specific factor. This way, the reasoning factor was extracted from the two spatial visualization tests on the predictor side, and the simple cardinal-axis rotation factor was extracted from the three complex skewed-axis rotation parcels on the criteria side.

Data preparation and analyses were conducted in R (R Core Team, 2021) using the lavaan package (Rosseel, 2012). Standard errors of all models were obtained using bootstrapping (10,000 resamples). All tests and analyses used $\alpha = .05$.

Models of mental rotation ability

In a first step, we compared two competing measurement models to test the incremental validity of a nested specific complex skewed-axis rotation factor as indicator of a change in cognitive processes when moving from simple cardinal-axis rotation to complex skewed-axis rotation tasks:

a) A one-factor model in which all parcels loaded on a single mental rotation factor
b) A nested-factor model consisting of a general factor on which both the simple cardinal-axis rotation parcels and the complex skewed-axis rotation parcels loaded, representing the simple cardinal-axis mental rotation processes required and used in both item types; and a specific factor on which only the complex skewed-axis rotation parcels loaded, representing unique cognitive processes that are required and used in these items, over and above those cognitive processes that are required and used in simple cardinal-axis rotation items.

In both models, loadings on the same factor, intercepts, and residual variances of corresponding parcels within the measurement model were constrained to be equal, to reduce model complexity, in particular since these item parameters were optimized to be comparable in data preparation. The relevance of the specific factor was then tested by evaluating its factor loadings and variance and by comparing the nested-factor model with the one-factor model in terms of model fit. As the models are nested, a comparison via χ^2 difference test was possible. Akaike information criterion (AIC), Bayesian information criterion (BIC), and sample-size adjusted BIC (aBIC) were also considered.

In a second step, mental rotation ability was predicted by the reasoning factor and by the spatial visualization factor. In doing so, we specified a nested-factor measurement model for the two predictors, with reasoning as the general factor and spatial visualization as the nested specific factor. This allowed investigating the specific contribution of spatial visualization in predicting mental rotation ability, over and above the contribution of reasoning.

4. Results

Hypothesis 1. Simple cardinal-axis rotation and complex skewed-axis rotation as distinct constructs.

Simple and complex rotation items did not differ in difficulty (two-sided paired t-test: $M_{\text{simple}} = 0.761$, $M_{\text{complex}} = 0.751$, $\Delta M = 0.011$, 95% CI = [−0.002;0.023], $t(371) = 1.632$, $p = .103$; d = −0.06).

Fit statistics of both the nested-factor model and the one-factor model were well within accepted fit criteria (Hu & Bentler, 1999; see Table 2). Model comparison showed that adding the spatial-visualization factor in the nested-factor model did not significantly improve model fit compared to the more parsimonious one-factor model ($\Delta\chi^2(1) = 3.512$, $p = .061$). Furthermore, the specific factor in the bifactor model was not adequately identified as it had weak standardized loadings ($\lambda = .23$), and its variance was not statistically significant ($s^2 = 0.002$, $SE = 0.001$, $p = .120$). This indicates that distinguishing between simple cardinal-axis rotation and complex skewed-axis rotation adds little value to the measurement model of spatial abilities for individual differences.

Hypothesis 2. Prediction of mental rotation factor(s) by reasoning (*g*) and spatial visualization.

To examine whether relations with reasoning and spatial visualization differ between simple cardinal-axis rotation and complex skewed-axis rotation we compared the two prediction models depicted in Fig. 2a and b.

Descriptive fit statistics of the prediction model with mental rotation as a single factor showed that it fitted the data well (see Table 3; structurally depicted in Fig. 2b). Although the χ^2 test was significant (which is frequently observed with large sample sizes; R. J. Vandenberg, 2006), the other fit criteria demonstrated that the model fitted the data well. Regression paths from both the general factor Reasoning and the specific factor Spatial Visualization were significant (Reasoning: $p < .001$; Visualization: $p = .022$; see Fig. 5).

Descriptive statistics of the prediction model with a nested-factor structure of mental rotation indicated good fit as well (see Table 3; structurally depicted in Fig. 2a), with the same caveat regarding the χ^2 test as with the one-factor mental rotation model. The test on fit difference of the nested-factor model and the one-factor model was not significant, indicating that both models fitted the data comparably well ($\Delta\chi^2(3) = 3.079$, $p = .380$). This indicated that estimating the additional parameters required for the more complex nested-factor model did not significantly improve model fit. Consequently, the more parsimonious one-factor model was accepted. Furthermore, the specific factor for complex skewed-axis rotation remained unidentified in the nested-factor model ($\lambda = .19$; $s^2 = 0.001$, $SE = 0.001$, $p = .262$). The lack of variance of the specific factor that can be accounted for in terms of reasoning and spatial visualization led to regression coefficients being virtually zero ($\beta_{\text{vis}} = .103$, $\beta_{\text{reas}} = .058$, $ps \geq .771$). Magnitudes of regression coefficients predicting the general factor for simple cardinal-axis rotation were similar in the nested-factor model and the one-factor model, indicating that the presence of the specific factor for complex skewed-axis rotation had negligible influence on the rest of the model.

These results led to the conclusion that the specific factor for complex rotation did not appear to have incremental value: Its inclusion did not significantly improve model fit and it captured a non-significant amount of variance that could not be accounted for either by reasoning or by spatial visualization. Therefore, we accepted the more parsimonious one-factor model with a general rotation factor depicted in Fig. 5 as the final model. This means that complex skewed-axis rotations were found not to constitute an ability that could be distinguished from simple cardinal-axis rotation ability. These findings can be reconciled with the notion that the processes resulting in individual differences in mental rotation are either the same in both item types or at least substantially correlated.

Furthermore, regression coefficients for the general rotation factor indicated that spatial visualization still had incremental predictive value beyond the expectably strong reasoning factor.

Table 2
Fit statistics of measurement models for mental rotation.

Model	χ^2	df	p	RMSEA	CFI	TLI	SRMR	AIC	BIC	aBIC
Nested-factor	20.97	21	.460	.00	1.00	1.00	.03	−1715.46	−1691.95	−1710.98
1 factor	24.49	22	.322	.02	1.00	1.00	.03	−1713.95	−1694.36	−1710.22

Note. aBIC = sample-size corrected BIC.

Table 3
Fit statistics of prediction models for mental rotation.

Model	χ^2	df	p	RMSEA	CFI	TLI	SRMR	AIC	BIC	aBIC
Nested-factor	83.97	51	.002	.04	.98	.98	.04	−3135.11	−3033.22	−3115.71
1 factor	87.05	54	.003	.04	.98	.98	.04	−3138.03	−3047.90	−3120.87

Note. Model names describe the measurement model used for mental rotation (as evaluated in Table 2) in the respective prediction model. aBIC = sample-size corrected BIC.

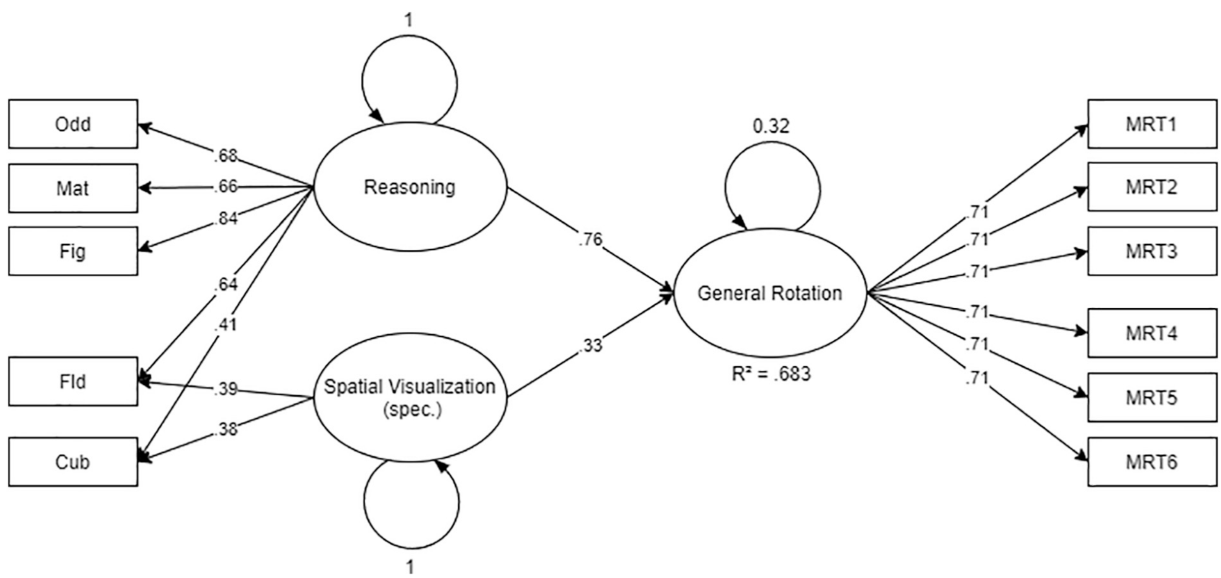

Fig. 5. Predictive model with general rotation as a single factor.
Note. Shown paths are standardized βs and are significant with α = .05.

5. Discussion

We investigated the popular Vandenberg and Kuse (S. G. Vandenberg & Kuse, 1978) mental rotation paradigm, which is frequently used for the assessment of individual differences in mental rotation ability. The test format builds on the R. N. Shepard and Metzler (1971) mental rotation test. However, most researchers compile their own tests using the Peters and Battista (2008) stimulus library. Consequently, test versions may differ in their processing requirements, and specific factors potentially affecting test performance are rarely considered in individual differences research.

Naturally, it could be hypothesized that similar factors would affect all forms of rotation test performance as those that were demonstrated for the R. N. Shepard and Metzler (1971) items (e.g., angular disparity, stimulus complexity or rotational complexity via rotational axes). This raises the question whether different test variants measure the same ability. A likely relevant factor in this respect is the number and type of rotational axes required. In typical mental rotation items (e.g., as summarized in S. Shepard & Metzler, 1988), stimuli need to be rotated around a single cardinal rotational axis. Usually, such a cardinal rotation axis can be easily derived either from the picture plane or from structural features of the stimulus (e.g., the edges of a cube figure). In such simple items, mental rotation around a cardinal axis may be the most central processing requirement. Conversely, more complex test items could be assumed to call for additional component processes, such as switching between different cardinal axes (i.e., the multiple-step rotation-by-dimension strategy) or searching for the one skewed rotational axis that allows for a one-step rotation to match the reference and target stimuli (i.e., complex skewed-axis rotation).

Previous studies testing the influence of rotational axes have revealed that complex skewed-axis rotations are more difficult, in terms of longer response times (Parsons, 1987); only high-spatial participants were able to use the skewed-axis rotation strategy (Just & Carpenter, 1985). However, the effect of rotational axes on the factor structure of mental rotation tests has not yet been investigated from an individual-differences perspective. This was the first goal of the present study.

To this end, we adopted a cognitive processes perspective and manipulated item characteristics accordingly (i.e., the required axes of rotation). Then, we tested differential effects of this manipulation using nested-factor analyses. This well-established method (Brunner, Nagy, & Wilhelm, 2012; Chen, West, & Sousa, 2006) allows to test whether – from an individual-differences perspective – component processes or abilities can be distinguished from each other. Furthermore, nested-factor analyses allow inspecting differential relations of the separated factors with other variables in the nomological network of relevant constructs (Messick, 1989).

Our results show that there is no specific ability for items requiring complex (skewed-axis) rotation, over and above general rotation ability. Specifically, a general factor of mental rotation ability accounted for individual differences in test performance in all rotation items. Conversely, there was no incremental factor capturing the specific requirements of complex skewed-axis rotation. This suggests that complex skewed-axis rotation does not constitute an ability on its own. This finding is of practical relevance for test construction. It suggests that the type of rotation (simple cardinal-axis rotation or complex skewed-axis rotation) does not affect the construct validity of the test. Consequently, our original concern that different test versions using different subsets of the Peters and Battista (2008) stimuli may differ in processing requirements was not supported. Conversely, it appears that the same ability is measured regardless of the required rotation requirements.

Furthermore, we did not find a meaningful mean difference in accuracy between simple cardinal-axis rotation and complex skewed-axis rotation items. This supports the supposition that distinguishing between both item types is not mandatory for the assessment of individual differences in mental rotation ability. This finding is well in line with Just and Carpenter's (1985) finding that there is no difference between rotating a cube once by 180° around one cardinal axis and rotating it twice, by 90° each time, around two different cardinal axes. Our results support that this finding also applies to more complex objects and to rotations around skewed axes not perpendicular to any feature of the rotated stimulus. (In Just and Carpenter's cubes, even skewed axes were always perpendicular to either faces, edges or corners of the cubes.)

However, our results did not confirm Parsons's (1987) important finding that rotations around skewed axes are more difficult, as reflected in slower reaction times in Parson's study. It is possible that this difference between simple cardinal-axis rotation and complex skewed-axis rotation is confined to reaction time data and has no effect on accuracy in untimed conditions. This difference could be due to differences in sensitivity or validity of the respective performance metric for cognitive processing requirements. These possibilities should be addressed in future research.

The second goal of the current study was to evaluate whether the Vandenberg and Kuse mental rotation test indeed necessitates a specific spatial-visualization ability over reasoning ability. This could be argued to be a critical threat to construct validity as previous research has shown (e.g., Hegarty (2018) that use of reasoning-based strategies can lead to better performance in these tests than spatial strategies alone. In order to test incremental validity, we fitted another nested-factor model to the tests we used as predictors. This allowed modeling a general factor of reasoning and an incremental factor capturing visualization-specific task requirements.

Next, rotation ability as assessed by the Vandenberg and Kuse test was regressed onto these two factors. The general factor of reasoning substantially predicted mental rotation ability. This was to be predicted, as reasoning tests usually reveal a substantial positive manifold (Carroll, 1993). This also holds for spatial ability tests in particular (Lohman, 1996). It is important to note, however, that the specific spatial-visualization factor incrementally predicted mental rotation ability. This finding confirms that there is a spatial-visualization component in the Vandenberg and Kuse paradigm, and that this component can be demonstrated over and above the substantial relation with general (reasoning) ability that has been shown for the Shepard and Metzler paradigm (see Mary Hegarty's work, e.g., Boone & Hegarty, 2017; Hegarty, 2018). This indicates that items are not exclusively solved by using analytical strategies such as feature matching, and spatial strategies such as mental rotation are used instead, at least to some extent.

5.1. Limitations

In the Vandenberg and Kuse paradigm, participants know the number of target stimuli per item. This led to two possible limitations in our study: First, responses to the two targets within each item were not fully independent of each other. This could have increased the estimated correlation between the two rotation factors. However, it could not account for their substantial relationship in the magnitude of identity. Nonetheless, this limitation could be alleviated in future research by presenting only one type of stimulus in each item (simple cardinal-axis or complex skewed-axis rotation).

A second limitation resulting from the number of targets per item being known is that it allowed solving an item by exclusion of distractors. If two distractors had been identified with sufficient certainty, participants did not need to look at the other two stimuli as these had to be the targets. This is especially problematic for the mental rotation test as a measure of spatial abilities as distractors (especially structure foils) can be detected via reasoning-based feature matching strategies more easily than targets, as it is easier to spot a single difference than to check the whole stimulus for identity. Therefore, distractors can be excluded by using both spatial strategies and analytical, reasoning-based strategies. Hegarty (2018) discussed this and questioned whether the mental rotation test following the Vandenberg and Kuse paradigm actually measures (exclusively) mental rotation ability: In her study she found that usage of some analytic strategies (namely what she called the "global shape strategy") also led to increased performance compared to most other reported strategies besides holistic rotation. Therefore, it may be possible that the test also measures the ability to find efficient analytical strategies which could account for its relations with reasoning abilities. Conversely, guessing based on an elimination strategy alone would be difficult to reconcile with the high internal consistencies (omega total $\omega_{simple} = .88$, $\omega_{complex} = .89$) observed in this study. Furthermore, the specific factor for spatial visualization ability still predicted mental rotation incrementally over and above reasoning, supporting that the test taps a spatial ability. However, we cannot rule out that participants also used analytical strategies (such as feature matching) or response elimination to supplement the spatial strategies.

There was no evidence of a specific factor for complex skewed-axis rotation. However, even if it had emerged it could not be unraveled whether participants had used the expected one-step rotation around the skewed axis for those items or had applied a two-step rotation-by-dimension strategy around two separate cardinal axes. Both the search process for the skewed axis required for the one-step rotation strategy and the process of switching between cardinal rotation axes in the two-step rotation strategy could have explained the emergence of a separate ability factor. In future research, self-reporting of solving strategy might give clues as to the strategies used by participants.

5.2. Conclusion

The present study contributes to a better understanding of mental rotation ability as an individual-differences construct. Using a psychometric approach applied to a large sample of participants, we showed that items in a mental rotation test following the Vandenberg and Kuse paradigm were accounted for by a single general factor. Conversely, there was no evidence of a specific processing ability required for the skewed-axis rotation items. Hence, mental rotation appears to be a unitary factor. Furthermore, we confirmed that the Vandenberg and Kuse paradigm measures spatial-visualization ability over and above reasoning ability. The implication for assessment purposes is that the Vandenberg and Kuse test measures spatial-visualization ability. Further, this ability can be assessed comparably regardless of whether simple cardinal-axis rotation or complex skewed-axis rotation items are used.

Acknowledgements

The preparation of this paper was supported by grant LE 645/18-1 from the German Research Foundation (DFG). The funding source had no further involvement in the study.

Appendix. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intell.2022.101626.

References

Arendasy, M. E., & Sommer, M. (2013). Reducing response elimination strategies enhances the construct validity of figural matrices. *Intelligence, 41*(4), 234–243. https://doi.org/10.1016/j.intell.2013.03.006

Boone, A. P., & Hegarty, M. (2017). Sex differences in mental rotation tasks: Not just in the mental rotation process! *Journal of Experimental Psychology. Learning, Memory, and Cognition, 43*(7), 1005–1019. https://doi.org/10.1037/xlm0000370

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80*(4), 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Bungart, M. (2018). *Entwicklung und Evaluation eines Manipulationsansatzes der Itemschwierigkeit in dem distraktorbasierten Format des Würfelkonstruktionstests [Development and evaluation of a manipulation approach to item difficulty in the distractor-based format of the cube construction task] [Unpublished bachelor's thesis]*. Saarbrücken: Universität des Saarlandes.

Burin, D. I., Delgado, A. R., & Prieto, G. (2000). Solution strategies and gender differences in spatial visualization tasks. *Psicológica, 21*(2), 275–286. https://www.redalyc.org/articulo.oa?id=16921204.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology, 7*(1), 20–43. https://doi.org/10.1016/0010-0285(75)90003-1

Doyle, R. A., & Voyer, D. (2018). Photographs of real human figures: Item types and persistent sex differences in mental rotation. *Quarterly Journal of Experimental Psychology (2006), 71*(11), 2411–2420. https://doi.org/10.1177/1747021817742079

Enders, C., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 430–457. https://doi.org/10.1207/s15328007sem0803_5

Flanagan, D. P., & Dixon, S. G. (2013). The Cattell-Horn-Carroll theory of cognitive abilities. In C. R. Reynolds, K. J. Vannest, & E. Fletcher-Janzen (Eds.), *Wiley online library. Encyclopedia of special education: A reference for the education of children adolescents and adults with disabilities and other exceptional individuals* (4th ed.). John Wiley & Sons, Inc.. https://doi.org/10.1002/9781118660584.ese0431

Geiser, C., Lehmann, W., & Eid, M. (2006). Separating "rotators" from "nonrotators" in the mental rotations test: A multigroup latent class analysis. *Multivariate Behavioral Research, 41*(3), 261–293. https://doi.org/10.1207/s15327906mbr4103_2

Glück, J., Machat, R., Jirasko, M., & Rollett, B. (2002). Training-related changes in solution strategy in a spatial test: An application of item response models. *Learning and Individual Differences, 13*(1), 1–22. https://doi.org/10.1016/S1041-6080(01)00042-5

Hausmann, M., Slabbekoorn, D., van Goozen, S. H. M., Cohen-Kettenis, P. T., & Güntürkün, O. (2000). Sex hormones affect spatial abilities during the menstrual cycle. *Behavioral Neuroscience, 114*(6), 1245–1250. https://doi.org/10.1037//0735-7044.114.6.1245

Hegarty, M. (2018). Ability and sex differences in spatial thinking: What does the mental rotation test really measure? *Psychonomic Bulletin & Review, 25*(3), 1212–1219. https://doi.org/10.3758/s13423-017-1347-z

Hegarty, M., Keehner, M., Khooshabeh, P., & Montello, D. R. (2009). How spatial abilities enhance, and are enhanced by, dental education. *Learning and Individual Differences, 19*, 61–70. https://doi.org/10.1016/j.lindif.2008.04.006

Heil, M., & Jansen-Osmann, P. (2008). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *Quarterly Journal of Experimental Psychology (2006), 61*(5), 683–689. https://doi.org/10.1080/17470210701822967

Heller, K. A., & Perleth, C. (2000). *KFT 4-12+ R - Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision [KFT 4-12+ R - Cognitive ability test for 4th to 12th grades, revision]. Beltz*.

Hirnstein, M., Bayer, U., & Hausmann, M. (2009). Sex-specific response strategies in mental rotation. *Learning and Individual Differences, 19*(2), 225–228. https://doi.org/10.1016/j.lindif.2008.11.006

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Judd, N., & Klingberg, T. (2021). Training spatial cognition enhances mathematical learning in a randomized study of 17,000 children. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01118-4. Advance online publication.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92*(2), 137–172. https://doi.org/10.1037/0033-295X.92.2.137

Khooshabeh, P., & Hegarty, M. (2010). Representations of shape during mental rotation. In Association for the Advancement of Artificial Intelligence (Ed.), *2010 AAAI spring symposium series* (pp. 15–20). AAAI Publications. https://www.aaai.org/ocs/index.php/SSS/SSS10/paper/viewPaper/1166.

Krüger, J. K., & Suchan, B. (2016). You should be the specialist! Weak mental rotation performance in aviation security screeners - reduced performance level in aviation security with no gender effect. *Frontiers in Psychology, 7*, 333. https://doi.org/10.3389/fpsyg.2016.00333

LimeSurvey GmbH. (2019). *LimeSurvey: An open source survey tool (version 3.20.2+191119) [computer software]*. Hamburg, Germany: Limesurvey GmbH. http://www.limesurvey.org.

Lohman, D. F. (1996). Spatial ability and g. In I. Dennis, & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 97–116). Lawrence Erlbaum Associates, Inc.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. https://doi.org/10.1037/met0000144

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *The American Council on Education/Macmillan series on higher education. Educational measurement* (3rd ed., pp. 13–103). American Council on Education.

Morin, A. J. S., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM), and bifactor-ESEM. In G. Tenenbaum, & R. C. Eklund (Eds.), *Handbook of sport psychology* (4th ed., pp. 1044–1073). Wiley. https://doi.org/10.1002/9781119568124.ch51.

Muto, H., & Nagai, M. (2020). Mental rotation of cubes with a snake face: The role of the human-body analogy revisited. *Visual Cognition, 28*(2), 106–111. https://doi.org/10.1080/13506285.2020.1727598

Parsons, L. M. (1987). Visual discrimination of abstract mirror-reflected three-dimensional objects at many orientations. *Perception & Psychophysics, 42*(1), 49–59. https://doi.org/10.3758/BF03211513

Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition, 57*(2), 176–184. https://doi.org/10.1016/j.bandc.2004.08.052

Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition, 66*(3), 260–264. https://doi.org/10.1016/j.bandc.2007.09.003

R Core Team. (2021). *R: A language and environment for statistical computing (version 4.1.0) [Computer software]*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology, 41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Shepard, S., & Metzler, D. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 3–11. https://doi.org/10.1037/0096-1523.14.1.3

Sorby, S. A., Veurink, N., & Streiner, S. (2018). Does spatial skills instruction improve STEM outcomes? The answer is 'yes'. *Learning and Individual Differences, 67*, 209–222. https://doi.org/10.1016/j.lindif.2018.09.001

Spearman, C. (1904). "general intelligence" - objectively determined and measured. *American Journal of Psychology, 15*, 201–293. https://psychclassics.yorku.ca/Spearman/.

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction, 17*(2), 219–234. https://doi.org/10.1016/j.learninstruc.2007.01.012

Stieff, M., Origenes, A., DeSutter, D., Lira, M., Banevicius, L., Tabang, D., & Cabel, G. (2018). Operational constraints on the mental rotation of STEM representations. *Journal of Educational Psychology, 110*(8), 1160–1174. https://doi.org/10.1037/edu0000258

Thissen, A., Koch, M., Becker, N., & Spinath, F. M. (2018). Construct your own response: The cube construction task as a novel format for the assessment of spatial ability. *European Journal of Psychological Assessment, 34*(5), 304–311. https://doi.org/10.1027/1015-5759/a000342

Thurstone, L. L. (1938). *Primary mental abilities*. The University of Chicago Press.

Vandenberg, R. J. (2006). Statistical and methodological myths and urban legends: Where, pray tell, did they get this idea? *Organizational Research Methods, 9*(2), 194–201. https://doi.org/10.1177/1094428105285506

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

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: A meta-analysis. *Psychonomic Bulletin & Review, 18*(2), 267–277. https://doi.org/10.3758/s13423-010-0042-0

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250–270. https://doi.org/10.1037/0033-2909.117.2.250

Weiß, R. H. (2006). *CFT 20-R Grundintelligenztest Skala 2 [CFT 20-R Base intelligence test scale 2]*. Hogrefe.

Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance, 24*(2), 397–412. https://doi.org/10.1037/0096-1523.24.2.397

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience, 20*(1), 1–19. https://doi.org/10.1162/jocn.2008.20013

Ziemek, T., Creem-Regehr, S., Thompson, W. L., & Whitaker, R. (2012). Evaluating the effectiveness of orientation indicators with an awareness of individual differences. *ACM Transactions on Applied Perception, 9*(2), 1–23. https://doi.org/10.1145/2207216.2207218